

分散環境における データベース利用支援の研究

江草由佳

図書館情報メディア研究科

筑波大学

2004年11月

A Study of Supporting to Use Databases in a Distributed Environment

Yuka Egusa

Graduate School of Library, Information and Media
Studies

University of Tsukuba

November 2004

要旨

本研究では、分散環境におけるデータベース利用支援の研究を行なった。現在、インターネットと World Wide Web (WWW) の普及にともない多種多様なデータベースサービスが提供されるようになり、場所、時間を問わず利用可能なデータベースは増え続けている。その上、これらの中には無料で提供されているものも多く、以前より、手軽に多くのデータベースを利用できるようになった。一方で、利用者による検索行動も多様化している。これは、利用者自身の多様化によるところが大きい。つまり、WWW 以前には専門家が担ってきた検索はエンドユーザの手に移り、児童から主婦・老人と、老若男女や職業を問わず、様々な種類のユーザが検索システムを利用するようになった。

本研究では、このように多種多様なデータベースが分散的に存在し、データベースの利用者が多様になったゆえにおこる問題の中でも、以下の2つに着目した。

1. 利用者のニーズに合ったデータベースを探し出し選択することの難しさ
2. データベースを使いこなすことの難しさ

利用者のニーズに合ったデータベースを探し出し、選択することができないと、検索結果が得られなかったり、検索する必要のないデータベースを検索し、検索に余計な時間がかかってしまったり、検索結果を検証する時間が増えてしまったり、多くのノイズが含まれる検索結果の中に必要な情報が埋もれたりする。そのため、適切なデータベースを探し出したり、選択することは、多種多様なデータベースが存在する場合必要なプロセスである。しかし、多種多様なデータベースが、分散的に構築され、日々データベースが変化する場合は、どのようなデータベースが存在し、利用者自身の検索におけるニーズ(利用のための時間や料金、対象とする分野、対象とするメディアなど) に合ったデータベースを把握することが難しいため、ニーズに合った適切なデータベースを選ぶことは難しい。

次に、データベースを使いこなすことの難しさについて説明する。データベースの機能には様々なものがある。例えば、フリーキーワードを使って全文検索、統制語を使った検索、タイトルや著者名などのフィールドを指定した検索、検索結果の並べ替え、などである。例えばあるフリーキーワードで検索して、ヒット件数が多かった場合、全ての検索結果を見るのは大変なため、タイトル

に絞って検索するなどがある。このように、データベースを検索し検索要求を満足させる結果を得るには様々な機能を使いこなす必要があるため、これらに習熟し使いこなすことは難しい。

本研究では、これらの問題を、検索ユーザインタフェースの統一と、データベースに関する情報の共有の2つのアプローチで解決できると考えた。

筆者が提案する検索ユーザインタフェースの統一とは、「ユーザが、使い慣れた1つの検索ユーザインタフェースから複数のデータベースを同様に検索できること」を指す。本研究で提案する「検索ユーザインタフェースの統一」では、ユーザは1つの検索ユーザインタフェースのみを習熟するだけでよく、その使い慣れた1つの検索インタフェースのみで複数のデータベースを検索でき、新たに利用したいデータベースであっても今まで利用していたデータベースの使い方が変わることなく利用できることを目指す。

筆者が提案するデータベースに関する情報の共有とは、「データベースに関する情報を、共同で入力し、共同で利用できること」を指す。本研究で提案する「データベースに関する情報の共有」では、個々人や各組織が得意な分野のデータベースに関する情報を収集・整理し相互に補完を可能にすること、共有された情報の中から適切なデータベースを選択できたり、データベースの使い方を知ることが可能になることを目指す。

本研究では、大きくわけて以下の3つの研究を行ない、この2つのアプローチの有効性を示した。

第1の研究は、Z39.50 検索クライアントの構築である。情報検索の相互運用を考慮した通信プロトコル Z39.50 に対応し、日本語を使って検索し、検索結果を表示でき、ネイティブに動く Z39.50 検索クライアントを開発することにより、利用者が統一的なインタフェースで、複数のデータベースを検索できることを示した。

第2の研究は、Z39.50 データベース選択支援環境である。この研究は WWW と Z39.50 を透過的に利用できるデータベース選択支援環境を提案し、その環境を実現する WWW-Z39.50 クライアントを開発した。この環境では Z39.50 データベースに関する情報を WWW 上で共有して利用できるだけでなく、利用者が自由にデータベースを組織化し、利用者同士でお互いに参照・利用できる。また、共有された情報であるデータベースリストの中から任意の一つをクリックすることによって、該当するデータベースの検索画面が起動し、直ちに検索を開始できる。

第3の研究は、データベースに関する情報を利用者同士で共同入力・共同利用できるシステムを開発した。この研究では、まず、データベース選択に必要な情報を基本情報、付加情報の2つに分け、21項目からなる基本情報を設計するとともに、この基本情報を入力・閲覧できるシステムを作成した。さらに、複数のデータベースをさまざまな観点でグループ化した情報やデータベース毎の利用方法などからなる付加情報についても、入力・閲覧できるようにした。また、データベース台帳総覧のデータ 3,614 件に適用し、その有効性を示した。

このように本研究では、分散的に存在する多種多様なデータベースを多様な利用者が利用する状況において、検索ユーザインタフェースの統一とデータベースに関する情報の共有というアプローチで研究してきた。

検索ユーザインタフェースを統一することによって、データベース毎にユーザインタフェースが変わることなく、利用者は複数のデータベースを同様に利用できる。このことは個々の利用者にとってよいだけでなく、データベースを提供する側にもよい。それは、今迄、新たなユーザインタフェースの習熟ができないために、利用しなかった利用者が利用するようになるため、自身の提供したデータベースを使ってもらえる機会が増えるからである。本研究では、第1の研究で、検索ユーザインタフェースの統一を実現し、その有効性を示した。

データベースに関する情報を共有することによって、日々変化するデータベースについての情報を個人や組織が個々に追うことなく、情報を共有している人々全員がそれぞれ得意分野を担当するなどして、相互に補完することができることは、利用者にとって有用である。また、多くのデータベースを提供することは図書館などでは普通になってきており、図書館の利用者のために、データベースに関する情報を個々の図書館が提供しなければならなくなっている。そのため、データベースに関する情報を作成するコストが大きくなったり、同じデータベースに関する情報を重複して作成するような状況になっている。そこで、図書館のようなデータベースに提供する側でも、個々の提供機関におけるデータベース情報作成のコストを低減し、効率よく情報を提供する点において、データベースに関する情報を共有する手法は有用である。本研究では、第2、第3の研究において、データベースに関する情報の共有を実現し、その有効性を示した。

検索ユーザインタフェースの統一とデータベースに関する情報の共有のそれぞれが可能になるだけでなく、これら2つを組みあわせれば、より効果的に多種多様なデータベースの利用支援になる。つまり、片方だけが可能であっても、検索できないことがあるからである。検索ユーザインタフェースの統一がされていれば、利用者はどのデータベースにも同様の使いかたで検索できるが、データベースの内容がわからなければ、効率的な検索が行えないし、また、逆にデータベースに関する情報の共有によって、データベースの内容がわかっているにもかかわらず、データベース毎に使い方を覚えなければ、検索ができない。例えば、そのデータベースが厳密な統制語で制御されているデータベースの場合、フリーキーワードで検索して、ヒットする結果が出ないことがある。この問題を解決するには、データベースの内容（データベースが厳密な統制語で制御されている）を知らなければ、検索できないなどである。本研究では、検索ユーザインタフェースとデータベースに関する情報の共有の両方を第2の研究において実現し、その有効性を示した。

現在、本研究を始めた当時より、ますます、多くの有用なデータベースが手軽に利用できるようになっており、例えば、筑波大学の附属図書館が筑波大学内に提供しているデータベースだけでも100を越している。また、雑誌・論文が電子的に提供されることも急激に増えており、データベースを検索して見つけた論文の全文が、図書館に行かずとも、読めるようになってきている。そのため、分散環境におけるデータベースの利用支援は、ますます重要になってきており、本研究の意義は大きい。

Abstract

I studied supporting to use databases in a distributed environment. Many more databases are available than in the past, because information resources, including databases accessed without restrictions of time or a place, have been increasing with the spread of the Internet and many of them are offered freely. Moreover, users' behaviors have been varied because of the variety in the database users. Before the spread of the WWW, most users of databases were specialists, but today most of the users are end users of various jobs and ages.

In this study, I focused following two problems caused by various databases and various database users:

1. It is difficult to find out and select databases which satisfy user's needs.
2. It is difficult to make full use of databases.

If a user can not find out and select databases which satisfy the user's needs, it causes some problems. 1) The user can not get the search result which the user wants to get. 2) It takes a long time to search many inappropriate databases. 3) It takes a long time to examine the search result because information which the user wants is buried in noisy search result. Therefore, the process to find out and select databases in the case which various databases are available is necessary.

For this process, it is necessary to know information about databases. But it is difficult to know this information because various databases are available on the world and user's needs vary (e.g. time and money to use, subject of database, available media).

Additionally, it is necessary to make full use many functions of databases. There are many functions: e.g. searching with free keyword in full-text search, searching with controlled term, searching with field limitation (author, title) and/or sorting search result. If a user searches a database using free keywords and the database returns many results, the user may need to search again with field limitation in order to narrow down the result. Therefore it is difficult to know many functions and make full use them.

I solved these problems using two approaches: unifying search user interface and sharing informa-

tion about databases.

My approach “unifying search user interface” means that users can search many databases with a single search user interface which they are familiar with. This approach allows users to search many databases with a search client and to use a new database with the client in the same way as other databases which they have used.

My approach “sharing information about databases” is one in which users are able to input and use information about databases with each other. This approach allows users to share information about various databases with each other. A user is able to find suitable databases for his/her information needs through the sharing information, to understand how to use the databases, and organize that information from his/her point of view.

My study is roughly divided into three parts. I showed the effectiveness of these two approaches through the study.

First, I developed a Z39.50 search client. This client is based on Z39.50 protocol for interoperability of information retrieval systems. By using the client, users are able to input Japanese search term and to display Japanese search result. I showed that users are able to search many databases with a single search user interface.

Second, I studied the environment that supported selection of Z39.50 databases. The environment allows users to share information of Z39.50 databases in WWW. I developed the WWW-Z39.50 client as a part of this environment. In this environment, the users can share information which other users organize by themselves. When a user clicks one of list of Z39.50 databases, which is shared information, the window for retrieval of the database will be opened for the user to start his/her retrieval immediately.

Third, I developed a system which users are able to input and browse information about databases. I classified information about databases into two categories. One is the fundamental information sufficient for users to select databases depending on their own needs. I designed 21 items as fundamental information and developed the system such that users are able to browse and add fundamental information. The second type of information is commentary information, which is comprised of, for example, users’ know-how and grouping information from various viewpoints. In my system users are able to browse and add commentary information. I have used 3,614 records from databases in Japan, derived from Database Daicho Soran, as a sample data set for my system. The system allows users to share information about databases, easily discover suitable databases and use them.

As I mentioned above, my approaches to support for various users to use various databases are “unifying search user interface” and “sharing information about databases”.

“Unifying search user interface” allows users to use multiple databases without changing the interface of each database. This benefits database providers as well as users of databases. The reason why

unifying interface benefits providers is the increasing opportunity to get their own databases used by a new user who have never used it because of unfamiliar interface with it. In my first study, I showed the unifying search user interface and the effectiveness of it.

Sharing information about databases benefits users because individual users or organizations do not need to get daily changing information and each user to share information contributes information about databases which the user is familiar with. Additionally, because a library often provides a lot of databases, each library has to provide more information about the databases for users.

Therefore, the cost of making information about databases has been increasing and information about same database has been made redundantly. “Sharing information about database” benefits providers of database service such as library in terms of providing effectively the information and decreasing the cost to make the information. In second and third studies, I showed the approach of sharing information about databases and the effectiveness of it.

Not only each approach, unifying search user interface and sharing information about databases, benefits supporting to use databases, but also the combination of these approaches benefits it more effectively. If only one approach is available, a user may be not able to search databases. In the case that only unifying search user interface is available, users are able to search all database in the same way. But users may not search databases effectively because they can not know information about databases. In the case of that only sharing information about databases is available, users are able to know information about databases. But users can not search the database without mastering how to use a database. For example, in the case of using a database which uses strict controlled term, if a user searches the database with free keywords, it returns inappropriate results. To solve this problem, it is necessary for users to know the information that the database uses strict controlled term. In second study, I showed the combination of these approaches and the effectiveness of it.

Many more useful databases are available today than at the time when I began this study. For example, databases provided by University of Tsukuba Library are over one hundred. Additionally, magazines and journals provided by a digital format through the Internet are increasing. People are able to read the full-text of a paper immediately which they find out by searching a database. Therefore, supporting to use databases in a distributed environment has been more important and this study is significant.

目次

第 1 章	序論	1
1.1	背景	1
1.2	データベースの定義	3
1.3	本研究でのアプローチ	4
1.3.1	検索ユーザインタフェースの統一	4
1.3.2	データベースに関する情報の共有	6
第 2 章	Z39.50 による日本語書誌データ検索システム	8
2.1	はじめに	8
2.2	Z39.50	9
2.2.1	ASN.1	10
2.2.2	BER	11
2.2.3	Request と Response	11
2.2.4	Initialize 機能 (接続機能)	12
2.2.5	Search 機能 (検索機能)	12
	検索式の構造	13
2.2.6	Present 機能 (返戻機能)	14
2.3	Z39.50 検索システム	16
2.3.1	システムの概要	16
2.3.2	日本語書誌データの処理	16
2.3.3	Z39.50-ULIS3 サーバ	17
2.3.4	Z39.50-ULIS3 クライアント	19
2.3.5	システムの実行例	22
2.3.6	開発システムの変遷	24
2.4	考察	27
2.4.1	システムの意義	27

2.4.2	Z39.50 検索システムの将来	28
2.5	まとめ	31
第 3 章	Z39.50 データベース選択支援環境	33
3.1	はじめに	33
3.2	データベース選択支援環境	34
3.3	WWW-Z39.50 クライアント	36
3.3.1	WWW-Z39.50 クライアントの概要	36
3.3.2	WWW ブラウザ	37
3.3.3	Z39.50 クライアント	38
3.3.4	実行例	39
3.3.5	開発環境	41
3.4	データベースリンク集の作成	41
3.5	考察	43
3.5.1	directory-of-servers との比較	43
3.5.2	Z39.50-WWW ゲートウェイとの比較	44
3.5.3	その他の関連研究	46
3.6	まとめ	47
第 4 章	分散環境におけるデータベースに関する情報の共有システム	48
4.1	はじめに	48
4.2	関連研究	49
4.3	データベースに関する情報の共有	50
4.3.1	データベースの基本情報	52
4.3.2	データベースの付加情報	52
4.3.3	サンプルデータの作成	55
4.4	データベースに関する情報共有システム	56
4.4.1	システムの構成	56
4.4.2	ブラウジング機能	58
	簡易表示	58
	詳細表示	59
	一覧表示	61
	ユーザページ表示	62
4.4.3	検索機能	62
4.4.4	基本情報の登録・更新機能	63

4.4.5	コメント機能	63
4.4.6	グループ化機能	63
4.4.7	ユーザ認証	66
4.5	考察	67
4.5.1	情報共有システム	67
4.5.2	基本情報の共同入力・共同利用	69
4.5.3	本システムの運用方法	70
4.6	まとめ	72
第 5 章	結論	73
	参考文献	76
	謝辞	83
	発表文献一覧	85

目次

2.1	Z39.50 の全体像	10
2.2	Z39.50	11
2.3	Z39.50 の PDU の最上位階層部分の ASN.1 での表現	12
2.4	BNF で表した検索式の構造	13
2.5	Japan/MARC の変換	17
2.6	検索処理	19
2.7	返戻処理	20
2.8	接続ウィンドウの機能	21
2.9	検索ウィンドウの機能	22
2.10	検索ウィンドウ：Z39.50-ULIS3 サーバ	23
2.11	検索ウィンドウ：図書館情報大学デジタル図書館システム Z39.50 サーバ	24
2.12	検索ウィンドウ：University of Wisconsin, Madison	25
3.1	データベース選択支援環境の全体像	35
3.2	システムの全体像	36
3.3	Z39.50 データベースを表す URL を用いた HTML の例	37
3.4	WWW ブラウザのシステム構成	38
3.5	WWW ブラウザ	39
3.6	Z39.50 クライアントの起動画面	40
3.7	ドキュメント表示例	41
3.8	複数データベース検索実行例	42
3.9	Z39.50 クライアントの検索画面	43
3.10	データベースリンク集の作成	44
3.11	WWW-Z39.50 ゲートウェイシステムとプロトコルとの関係	45
3.12	本システムのプロトコルとの関係	45
4.1	データベースに関する情報の共有イメージ	51

4.2	データ例	54
4.3	システムの構成図	58
4.4	最新基本情報簡易表示 (初期画面)	59
4.5	基本情報詳細表示	60
4.6	主題一覧	61
4.7	ユーザのページ	62
4.8	基本情報詳細表示における操作例 (1)	64
4.9	基本情報詳細表示における操作例 (2)	65
4.10	ユーザ自身のページ (新規グループ作成)	66
4.11	ユーザ自身のページ (新規グループ作成後)	67
4.12	グループ編集	68
4.13	基本情報詳細表示における操作例 (3))	69

表目次

2.1	bib-1(抜粋)	15
2.2	Bib-1 と Japan/MARC の対応表	18
2.3	開発システムの相違点	26
4.1	基本情報の項目説明	53
4.2	基本情報とデータベース台帳の項目との対応表	57

第 1 章

序論

1.1 背景

現在、インターネットと World Wide Web (WWW) の普及にともない多種多様なデータベースサービスが提供されるようになっている。例えば、論文中のタイトルや著者アブストラクトなどが検索できる書誌データベース、図書館が所蔵している書籍を検索できる OPAC、新聞記事を記事全体や見出しで検索できる新聞データベース、ある特定の日や年間の日照時間・気温などの気象情報にみられるような数値的な情報が検索できるファクトデータベース、WWW で提供されているドキュメント全体をキーワード検索できるサーチエンジン、ある WWW サイト内のみドキュメントをキーワード検索できるサイト内検索、料理のレシピを材料などで検索できるデータベース、薬の使いかたや副作用を検索できるデータベースなどである。これらの中には無償で利用できたり、利用のための場所、時間の制約がないものも多く、以前より、手軽に多くのデータベースを利用できるようになった。

また、WWW 以前には専門家が担ってきた検索はエンドユーザの手に移り、児童から主婦・老人と、老若男女や職業を問わず、様々な種類のユーザが検索システムを利用するようになり、利用者は以前よりも多様化した。それにともない、利用者の検索行動も多様化している。例えば、以下のようなパターンがある。

1. 主婦が毎日の献立の参考に料理のレシピを検索する。
2. 患者が自分に処方された薬の保存方法を検索する。
3. ビジネスマンが現在話題になっている時事情報の詳細を検索する。
4. 視聴者が TV で放映されている映画の登場人物が他にどの映画作品に出ているか検索する。
5. 消費者が現在購入予定の家電の市場価格を検索する。
6. 消費者が現在購入予定の化粧品の評判（口コミ情報）を検索する。
7. 研究者が自身の研究課題と類似した既存の研究がないかどうかを検索する。

8. 企業の特許部門の人が既存の特許の中に自身の出願と重複した部分がないかどうか検索する

本研究では、このように多種多様なデータベースが分散的に存在し、データベースの利用者が多様になったゆえにおこる問題の中でも、以下の2つに着目した。

1. 利用者のニーズに合ったデータベースを探し出し選択することの難しさ
2. データベースを使いこなすことの難しさ

利用者のニーズに合ったデータベースを探し出し、選択することができないと、検索結果が得られなかったり、検索する必要のないデータベースを検索し、検索に余計な時間がかかってしまったり、検索結果を検証する時間が増えてしまったり、多くのノイズが含まれる検索結果の中に必要な情報が埋もれたりする。

中でも特に問題なのは、以下のように、適切なデータベースを探し出さないと、検索結果が得られない場合である。田澤 [1] によれば、「過去、直近 20 年の間で、北海道北見市で一番『年間日照時間』が長かったのは、西暦何年か」という検索要求の場合、単に Google などの Web 全体の Web ページ全般のサーチエンジンで「『日照率』『年間日照時間』『北見市』などのキーワードで検索しても、回答はなかなか見つからず、「『北見は日照時間が日本有数である』という、数字を伴わない情報が中心」である。気象庁の「電子閲覧室」という気象データを扱ったデータベースを検索しなければ回答にたどり着かない。それは、「電子閲覧室」データベースの検索結果である年間日照時間のように、動的に生成される Web ページを Web 全体を扱っている伝統的なサーチエンジンが、検索サービスの対象外としている [2] ためである。そこで、この検索要求を満たすには、サーチエンジンのその特性を知り、かつ、気象データを扱っているデータベースを見つけ、検索できなければならない。

そのため、適切なデータベースを探し出したり、選択することは、多種多様なデータベースが存在する場合必要なプロセスである。しかし、多種多様のデータベースの中から、利用者のニーズに合った適切なデータベースを選ぶことは難しい。その理由を次に述べる。

利用者の検索要求、検索にかかる時間や金額や手間、検索結果に求める網羅性・正確性の程度は様々である。例えば、日常のちょっとした疑問「今日の晩御飯のチンジャオロースの味付けには、どんな調味料が必要？」では、いくつかのレシピに載っている代表的な味付け方法を眺められればよく、無償で検索でき、短時間で結果を得られることが優先される場合がある。また、一方で、研究者による「これから自分がしようと思っているある研究がすでになされているかどうか」のような関連研究の検索では、利用者はとにかくもれのない正確な検索結果を得ることを優先するため、検索が有料となる、図書館からしか検索できないデータベースである、結果を得るまでに 1 週間くらいかかるといった検索に要するコストや時間にはこだわらない場合も多い。

このような様々なニーズに合ったデータベースを選ぶためには、どのようなデータベースが存在するかを知る必要がある。例えば、利用料金が無料・有料のもの、利用場所が限定されている・さ

れていないもの、対象分野、対象メディアなどである。

しかし、多種多様なデータベースが、分散的に構築され、日々データベースが変化する場合、どのようなデータベースが存在し、自身の検索におけるニーズ（利用のための時間や料金、対象とする分野、対象とするメディアなど）に合ったデータベースを把握することが難しいため、ニーズに合った適切なデータベースを選ぶことは難しい。

次に、データベースを使いこなすことの難しさについて説明する。データベースの機能には様々なものがある。例えば、フリーキーワードを使って全文検索、統制語を使った検索、タイトルや著者名などのフィールドを指定した検索、検索結果の並べ替え、などである。例えばあるフリーキーワードで検索して、ヒット件数が多かった場合、全ての検索結果を見るのは大変なため、タイトルに絞って検索するなどがある。このように、データベースを検索し検索要求を満足させる結果を得るには様々な機能を使いこなす必要がある。

1.2 データベースの定義

本節ではまず、従来なされてきたデータベースの定義をいくつか紹介し、それらをふまえた本研究におけるデータベースの定義を述べる。

C.J.Date は、データベースとは「ある特定組織のアプリケーション・システムによって使用される永続的データの集まりである」[3, p.10] と定義している。また、Garcia-Molina らはデータベースを「一定の期間に渡り、保持される情報の集まりである」[4, p.2] と定義している。

一方、2003 年版データベース白書 [5] ではデータベースを「1:特定のテーマに基づいてデータを体系的に整理または整理のつく状態で保存したもの、2:データの集まりの中から必要なものだけを指定して情報としての部分データとして取り出せるもの、3:パソコンや携帯情報端末などのコンピュータ機能を備えている情報端末機器で検索可能な形態になっているもの」の 3 基準を満たすデータの集まりと定義しており、実際に、利用者が利用するデータベースを想定した定義になっている。

また、ウェブ上の各種データベースへのリンクを提供するゲートウェイ・サービス「国立国会図書館データベース・ナビゲーション・サービス」では、「対象とするデータベース」[6] として、「Web でその情報が開示されており、一定のフォーマットに基づき標準化された内容の複数のレコードが、検索やブラウジングを目的に組織化されており、データベース内では、全レコードが共通の要素で関係付けられているもの」とし、さらに、「1:利用者による何らかのアクション（検索語投入、ボタン押下等）によって、背後のデータベースの検索が実行され、検索結果が動的に HTML として生成・表示される。2: 静的な情報源であっても、階層ディレクトリ（ディレクトリによる階層型リンク構造によるデータ集合）や簡易検索などのナビゲーション機能が提供されている。3: FTP 等によるダウンロードまたは媒体による、データの一括送付を提供している。」のいずれかの機能を備えていると定義している。これは、とりわけ Web 上のデータベース提供の実体に沿うように定義がなされたものと思われる。

また、本研究においてそのシステム構築を行った Z39.50 では、データベースを「関連する情報を含む情報単位の集まり」[7, p.4] と定義している。

以上の定義を踏まえて、本研究では、データベースを以下のように定義する。データベースとは、電子的に蓄積された複数のレコードであるデータの集まり、または、それを操作するシステムを含むものを指す。

「データの集まり」をデータベースとして定義し、「それを操作するシステム」はデータベースマネジメントシステム (DBMS) として定義するという考えかたもあるが、本研究で「データの集まり」だけでなく、「それを操作するシステム」をもデータベースの定義に含めた理由を以下に挙げる。

利用者が実際にデータベースを使う場合に問題となるのは、データの集まりの内容だけでなく、それら进行操作・検索するシステムの機能にあることも多い。そのため、データベースの利用支援といった場合には、データの集まりだけでなく、データの集まりとそれら进行操作・検索するシステムの機能をセットで扱う必要があるからである。例えば、同じデータの集まりが異なるシステムでサービスされている場合、利用者としては、そのデータの集まりをどのような機能で操作・検索できるかが重要となる。

この定義にもとづいたデータベースには以下のようなものがあげられる。Japan/MARC (書誌データの集まり)、サーチエンジン Google (Web ページのデータの集まりとそれを検索するシステム) など。

1.3 本研究でのアプローチ

本研究ではこの 2 つの問題にどのようなアプローチで取り組んだかについて述べる。本研究では、これらの問題を、以下の 2 つのアプローチで解決できると考えた。

- 検索ユーザインタフェースの統一
- データベースに関する情報の共有

1.3.1 検索ユーザインタフェースの統一

データベースそれぞれに利用方法やユーザインタフェースが異なるため、それぞれのデータベースの使いかたを知らなければ、検索要求を満足する結果は得られない。

データベースの利用者が、サーチャーや図書館司書など、情報検索というタスクそのものの優先度が高い職業であれば、自らの経験や知識をもとに検索することができる。もしデータベースの使いかたを知らなかったとしても、利用方法を調べたり、データベース活用の為の研修会に出たりなどして、データベースを利用・活用するために労力をかけて、データベースを利用できるが、現在

のようにデータベースが流動的であると、専門家であっても、個々の検索ユーザインタフェースの習得は困難である。

一般の利用者の場合は、たとえアクセス可能なデータベースが多種多様にあったとしても、検索自身が主要な仕事でないので、新たなデータベースを使うために、新たにその利用方法を習得することまでは避けたい心理が働き、使い方を知っているデータベースのみを利用することになってしまい、せっかく有用な多くのデータベースがあっても実際には利用できなくなっているという問題がある。また、例えばフリーキーワードを使った簡単な検索はできるけれども、収録期間を限定した検索はできないなど、データベースの詳細な使いかたがわからないために、検索要求にあった結果が得られないなどの問題もある。

そこで、本研究では、検索ユーザインタフェースを統一することでこの問題が解決できると考えた。筆者が提案する検索ユーザインタフェースの統一とは、「ユーザが、使い慣れた1つの検索ユーザインタフェースから複数のデータベースを同様に検索できること」を指す。本研究で提案する「検索ユーザインタフェースの統一」では、ユーザは1つの検索ユーザインタフェースのみを習熟するだけでよく、その使い慣れた1つの検索インタフェースのみで複数のデータベースを検索でき、新たに利用したいデータベースであっても今まで利用していたデータベースの使い方が変わることなく利用できることを目指す。なお、「ユーザインタフェースの統一」と言った場合には、デスクトップ環境の look&feel といった面におけるユーザインタフェース統一を図るためのガイドライン [8][9] を整備してその環境上で動作するアプリケーションソフトウェアのユーザインタフェース統一を目指す動きもある。しかし、上述した通り、本研究における「検索ユーザインタフェースの統一」はこのような特定の look&feel に基づくユーザインタフェース・ガイドライン整備やアプリケーションソフトウェア固有のユーザインタフェースについてはその範囲としない。

そこで本研究では、異なる複数のデータベースを共通の検索ユーザインタフェースで検索できる Z39.50 プロトコルに着目した。

Z39.50 プロトコル検索システムの利用者の観点からの利点は、データベース毎に検索ユーザインタフェースが変わることがなく、一つのクライアント(=検索ユーザインタフェース)を用いて、複数のデータベースを同様に検索できることにある。使いなれたクライアントが一つあれば、新たにデータベースが増えたとしても、利用者は一からその利用方法(検索ユーザインタフェース)を覚える必要はない。

しかし、筆者が研究をはじめた当時、Z39.50 における日本語をはじめとする非欧米言語の処理については未発達の状態であり、日本語に対応した検索クライアントは見あたらなかった。そこで世界中の Z39.50 サーバを日本語ユーザインタフェースを用いて検索可能な Z39.50 クライアントを構築し、情報知識学会に論文 [10] として発表した。なお、上記論文における筆者の担当箇所は、Z39.50 クライアントの設計と開発である。このシステムについての詳細は2章で述べる。

1.3.2 データベースに関する情報の共有

利用者が、利用可能なデータベースを把握し、適切に選択するためには、どのようなデータベースが存在するか、それぞれのデータベースはどんな分野を扱っているか、どのようなレコード（例：書誌、要約）を持っているかなどを知る必要がある。そのため、利用者は、データベースディレクトリや、分野ごとにまとめられたノウハウ本、利用可能な図書館が作成したデータベースリストから、これらデータベースに関する情報を得ている。

しかし、現在のように分散的に多種多様なデータベースができては消え、変化している場合、検索システムをそれぞれの機関が作成しているために、データベースに関する情報をまとめて得ることが難しく、また情報の変化が速い。そのため、これらのデータベースに関する情報を個々人や各組織が、独立して収集・整理することは困難である。そこで、データベースに関する情報を共有する、つまり、共同で入力し、共同で利用することにより、個々人や各組織が得意な分野を収集・整理し、長所を生かしつつ、短所を補うことができると考えた。

1.1 節で述べたように、データベース利用者の多様化と、多様な検索行動におけるニーズによって、データベースに関する情報の作成・収集・整理へのニーズもまた多様化している。

データベース利用者の観点やデータベース利用のニーズによって必要・有用な情報が異なるからである。同じ1つのデータベースについての情報の作成であれば、主婦であれば無償で利用できることが重要であるかもしれないし、研究者であればデータベースの扱っているデータの正確性や網羅性が重要であるかもしれない。また、複数のデータベースに関する情報を収集・整理する場合であれば、大学の学生であれば所属している大学の附属図書館で利用可能な複数のデータベースのデータベースのリストがまとめてあればよいし、化学の研究者であれば化学の論文や解析データを扱ったデータベースのリストがまとめてあればよい。また、データベースの提供者であれば自身の提供するデータベースの良い部分を強調し他の似たデータベースに対して劣る部分については伏せたいかもしれないし、逆にデータベースの利用者であれば複数のデータベースの利用料金や機能の比較一覧表が見たいかもしれない。

そこで、本研究ではデータベース作成者や提供者だけでなく、データベースの利用者を含めた多様な個々人や組織が、個々の観点やニーズに対応したデータベースに関する情報を共同で入力し、利用することが必要と考えた。

本研究では、このデータベースに関する情報の共有についての研究を行なった。これらの成果は、特に Z39.50 データベースの選択について着目した研究として、情報知識学会に論文 [11] として発表した。このシステムの詳細については3章で述べる。

また、Z39.50 データベースだけでなく、全てのデータベースを対象とし、そして、データベースに関する情報をインターネット上のコミュニティを形成することで共有することについて着目した研究として、国際会議に口頭発表 [12] として発表し、また情報知識学会に論文 [13] として発表し

た。このシステムの詳細については 4 章で述べる。

第 2 章

Z39.50 による日本語書誌データ検索システム

2.1 はじめに

Z39.50[14] はクライアント・サーバ環境において、クライアントとサーバが通信する際のデータ構造やデータ変換の処理方式を規定したプロトコルである。すなわち、システム内部でのデータ構造や検索処理がシステムごとに異なっているとしても、通信する際に同じ構造と処理方式にしておけば、それぞれのシステム構成を維持したまま互いのシステムを利用できる。

1960 年代後半からオンライン検索システムの開発が始まり、1972 年には DIALOG や ORBIT などがデータベースベンタとして多くの商用データベースを抱え、オンラインサービスを開始した。時同じくしてアメリカ議会図書館、OCLC、RLIN などが機械可読目録 (MARC) のデータベースを構築し、書誌ユーティリティとしての機能を果たし始めた。これらのシステムは公衆回線を通じて端末からホストコンピュータにアクセスする中央集中型システムであり、この時代はこの形態のシステムが情報を得る上で最も効率がよかった。

しかしながら、中央集中型システムと言えども全ての情報が集中しているわけではないため、様々な情報要求を満たすためには複数のシステムを利用せざるを得なかった。ところが、ここで問題となったのはコストと使い勝手であった。すなわち、これらのデータベースを使用するにはそれぞれ高額の使用料を支払わなければならない、そのうえ、複数のデータベースの使用料に見合う情報は必ずしも得られなかった。特に書誌ユーティリティにおいては重複データが数多くあり、情報を得る上で効率が悪く、使い勝手の面でもシステムごとに接続手順や検索コマンドに違いがあり、利用者の検索要求を十分満たせないという状況であった。

こういう状況の下、1979 年に主要な書誌ユーティリティ間で目録データを交換し、互いのデータを透過的に検索できることを目指したプロジェクトが開始され、その後、OSI の規格と平行して ANSI の構成団体である NISO で議論され、1988 年に標準情報検索プロトコル Z39.50 として規格

が制定された [15][16]。

1990 年代に入り、TCP/IP をベースとするインターネットの普及にあわせ、Z39.50 の規格は OSI ベースの 1992 年の Version2 から 1995 年の Version3 に発展した。この時期に、アメリカ議会図書館、OCLC をはじめ、アメリカ、ヨーロッパで図書館システムを中心に数多くの Z39.50 検索システムが実装されるようになり [17][18][19]、1999 年 6 月現在、Index Data 社が作成しているサーバリスト [20] に登録されているだけでも 228 にのぼる。

日本でも欧米の動きに呼応するように、Z39.50 に関連したシステムが構築されるようになってきた。既存のシステムを使ったゲートウェイサービスとしては、OCLC の SiteSearch を使った紀伊国屋のサービス [21]、海外の Z39.50 サーバへのゲートウェイ機能を持つシステムとしては、丸善の MARUZEN Z39.50 Gateway-Cat[22]、科学技術振興事業団 (JST) の JOIS-withSTN[23] がある。図書館情報大学デジタル図書館 [24]、早稲田大学図書館の WINE[25]、東京工業大学の電子図書館システムには Z39.50 の検索システムが含まれており [26][27]、学術情報センターの電子図書館システム NACSIS-ELS は、クライアントとサーバの間のアクセスに Z39.50 を使用している [28][29]。また、地球観測衛星データシステムに Z39.50 を応用した例 [30] や、英語の書誌データを対象にした Z39.50 サーバの構築例もある [31]。また、石田による Japan/MARC を対象とした日本語検索のための仕様策定への提案も行なわれている [32]。

図書館情報大学 (ULIS) では 1995 年以来、日本語書誌データを対象に Z39.50 サーバのシステム構築実験を続け、1996 年に安斎によって WWW ブラウザから検索できるシステムが完成した [33]。このシステムを Z39.50-ULIS1 と呼ぶ。その後、我々はシステム構成を見直し、1998 年にクライアントシステムも含む Z30.50 の基本機能を完全に備えた Z39.50-ULIS2 を開発した [34]。引き続き、クライアント機能を改良するとともに、大規模データに耐えられるシステムにするため、Z39.50-ULIS2 のサーバ部分を全面的に作り直した。本論文はこの新しく開発した Z39.50-ULIS3 について報告する。Z39.50-ULIS3 はより実験システムとしての性格を強め、拡張性と移植性を高めた。

2.2 Z39.50

Z39.50 は検索クライアント検索サーバ間のやり取りを標準化しているため、Z39.50 を用いた検索クライアントを使うことにより、ユーザは固有のインタフェースやシステムの違いを意識せずに Z39.50 を用いた検索サーバにアクセスし、検索できる。また、Z39.50 はセッションを維持したままの検索を約束しているので、履歴を利用した検索もできる。この検索はステートフルな検索と言われる。

図 2.1 は Z39.50 検索システムのイメージを表している。左にクライアント、右にサーバがあり、その間のやりとりを Z39.50 で行なう。クライアントは、特定のサーバだけでなく、Z39.50 を用いたサーバであればどのサーバでもアクセスでき、また、サーバも特定のクライアントからだけでは

なく、Z39.50 を用いたクライアントであればどのクライアントからでもアクセスできる。

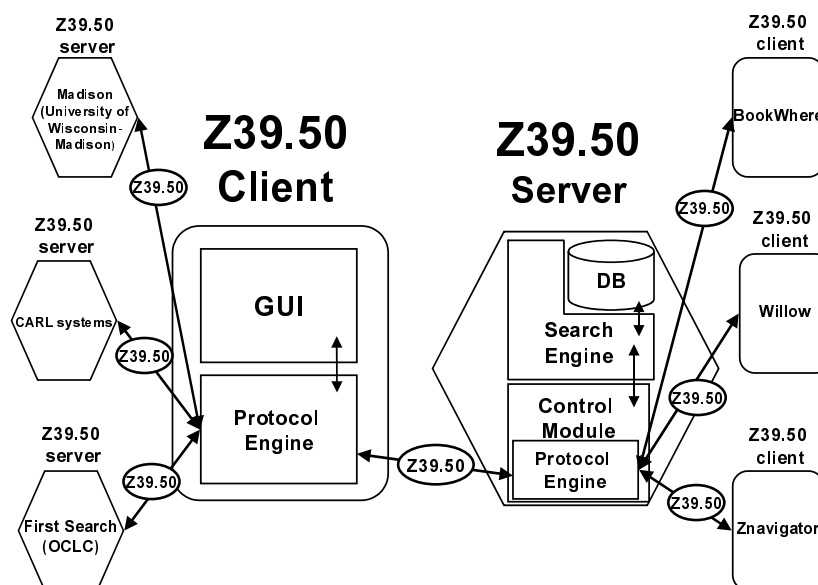


図 2.1 Z39.50 の全体像

Z39.50 は、検索質問、検索結果、課金、認証などクライアント、サーバ間のやりとりのみを標準化しており、クライアント、サーバはそれぞれ自由に構築できる。

Z39.50 はクライアントサーバ間でやり取りする値、つまり、接続の時に必要な情報である IP アドレス、検索の時に必要な検索式の構造、検索結果を返す時に必要な検索結果の構造、こういったものすべてを PDU(Protocol Data Unit:プロトコルデータ要素) として転送する。PDU は機種やアプリケーションに依存せずにデータが記述できる ASN.1 を用いて記述し、BER で符合化する (図 2.2)。

2.2.1 ASN.1

Z39.50 では、機種に依存しない通信を可能とするために、ASN.1 を使用している。ASN.1 (Abstract Syntax Notation.1 : 抽象化構文記法 1) とは ISO 標準の「PDU のデータ構造およびその PDU を転送する場合のオクテット列の形式を規定する言語」であり、「計算機に独立なデータ表現手法」[35] である。

図 2.3 は Z39.50 の PDU の最上位階層部分の ASN.1 での表現である。

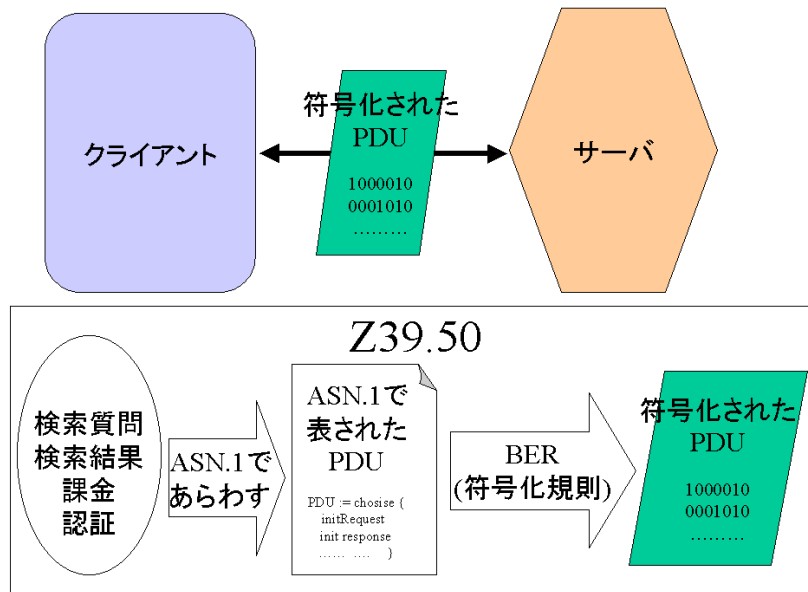


図 2.2 Z39.50

2.2.2 BER

ASN.1 は抽象的なデータ構造を表現する言語なので、ASN.1 で記述された PDU を、ネットワークで転送するためには、バイナリデータに変換する必要がある。BER (Basic Encoding Rules : 基本符合化規則) は ASN.1 で規定された構造のデータをバイナリデータに変換する符合化規則である。

2.2.3 Request と Response

Z39.50 では、クライアントからの要求を Request、サーバのクライアントの対する結果を Response といい、クライアントから Request が出されると、サーバからは必ず Response を返す。クライアントから出される要求である Request とサーバから出される Response は、Initialize, Search, Present, DeleteResultSet, AccessControl, ResourceControl, ResourceReport, Scan, Sort, Segment, ExtendedServices, Close の 12 あり、最も基本的なものは Initialize , Search , Present の 3 つである。

```

PDU ::= CHOICE{
  InitRequest           [20] IMPLICIT InitializeRequest,
  InitResponse          [21] IMPLICIT InitializeResponse,
  searchRequest         [22] IMPLICIT SearchRequest,
  searchResponse        [23] IMPLICIT SearchResponse,
  presentRequest        [24] IMPLICIT PresentRequest,
  presentResponse       [25] IMPLICIT PresentResponse,
  deleteResultSetRequest [26] IMPLICIT DeleteResultSetRequest,
  deleteResultSetResponse [27] IMPLICIT DeleteResultSetResponse,
  accessControlRequest  [28] IMPLICIT AccessControlRequest,
  accessControlResponse [29] IMPLICIT AccessControlResponse,
  resourceControlRequest [30] IMPLICIT ResourceControlRequest,
  resourceControlResponse [31] IMPLICIT ResourceControlResponse,
  triggerResourceControlRequest [32] IMPLICIT TriggerResourceControlRequest,
  resourceReportRequest [33] IMPLICIT ResourceReportRequest,
  resourceReportResponse [34] IMPLICIT ResourceReportResponse,
  scanRequest           [35] IMPLICIT ScanRequest,
  scanResponse          [36] IMPLICIT ScanResponse,
  -- [37] through [42] reserved
  sortRequest           [43] IMPLICIT SortRequest,
  sortResponse          [44] IMPLICIT SortResponse,
  segmentRequest        [45] IMPLICIT Segment,
  extendedServicesRequest [46] IMPLICIT ExtendedServicesRequest,
  extendedServicesResponse [47] IMPLICIT ExtendedServicesResponse,
  close                 [48] IMPLICIT Close}

```

図 2.3 Z39.50 の PDU の最上位階層部分の ASN.1 での表現

2.2.4 Initialize 機能 (接続機能)

クライアントは接続するための情報である接続要求 (Initialize Request) を出し、サーバはクライアントからの接続要求を受けるか拒否するかの情報、接続結果 (Initialize Response) を返す。Initialize Request には、プロトコルバージョン、サポートする機能、クライアントの情報などがある。Initialize Response にも同様に、プロトコルバージョン、サポートする機能、サーバの情報などがある。

2.2.5 Search 機能 (検索機能)

クライアントが、検索質問、検索質問の種類、データベース名などの検索要求 (Search Request) を出すと、サーバは検索結果 (Search Response) を返す。検索結果には、ヒットした件数などが含まれる。

検索式の構造

検索式の種類には、Type-0,Type-1,Type-2,Type-100,Type-101 が用意されており、一般的には Type-1 Query が使われる。Type-1 は逆ポーランド記法 (RPN Query) で記述されたツリー構造の検索質問である。図 2.4 は検索式の構造を BNF で表したものである [14]。

RPN-Query ::= Argument	
Argument + Argument + Operator	
Argument	::= Operand RPN-Query
operand	::= AttributeList + Term
ResultSetId Restriction	
Restriction	::= ResultSetId + AttributeList
operator	::= AND OR AND-NOT Prox
The notation above is used as follows:	
::=	means "is defined as"
	means "or"
+	means "followed by", and + has precedence over (i.e., + is evaluated before).

図 2.4 BNF で表した検索式の構造

検索質問には、検索語の属性を表す Attribute List と Attribute Set ID が付加される。Attribute List + Attribute Set ID は書名、著者名、前方一致、後方一致、などを示すアクセスポイントを示す。

- Attribute
検索語の属性を表す。Attribute List と、Attribute Set Id から構成される。
- Attribute Set
Attribute がどの属性に属しているのかを表す (書誌、全文など)。
- Attribute set ID
Attribute Set に Z39.50 が一意につけた ID で、書誌情報用に作られた bib-1 が通常使われる。
- Attribute Type
整数値で表される。Attribute set ID bib-1 では Attribute Type は 表 2.1 のように整数値が割り当てられている。
- Attribute Value
整数値で表される。Attribute set ID bib-1 では Attribute Value は 表 2.1 のように整数値が割り当てられている。
- Attribute Element

Attribute Type、Attribute Value の組合せで表される。

例えば、Attribute set ID bib-1 では以下になる。

- 1-1 Use Personal name(個人名)
- 1-4 Use Title(タイトル)
- 4-1 Structure word(ワード)
- 4-2 Structure phrase(フレーズ)
- 5-1 Truncation right Truncation (左方向のトランケーション)
- Attribute List
Attribute Element の集合で表される。
- Result Set ID
検索結果集合の ID を表す。

例をあげると、「書名に Library という語 がつく本を探したい」という場合、「(1-4 , 4-2)“Library” ,bib-1」となる。1-4 は書名を、4-2 は 検索語が語であることを表している。1-4, 4-2 は Attribute List、1-4 や 4-2 は Attribute Element、1-4 の 1 や 4-2 の 4 は、Attribute Type 1-4 の 4 や 4-2 の 2 は、Attribute Element を表している。bib-1 は Attribute set ID を表している。

2.2.6 Present 機能 (返戻機能)

検索した結果の書誌レコードを得る機能であり、search and retrieve の retrieve にあたる。クライアントは検索した結果の書誌レコードを要求する返戻要求 (Present Request) を出す。サーバは要求されたレコード返戻結果 (Present Response) を返す。一般的には、検索するとヒットした件数だけでなく、書誌事項も同時に表示するが、Z39.50 では通常、検索クライアントが検索要求を出すと、検索サーバはヒットした件数などを返し、書誌レコードは返さない。検索クライアントが検索結果の書誌レコードを得るには、検索要求とはまた別に返戻要求を検索サーバに出さなければならない。検索結果の形式にはプレーンテキストである SUTRS や、従来から使われている USMARC 形式がある。エレメントセットネームには簡略形式を表す B や、詳細形式を表す F がある。

表 2.1 bib-1(拔粹)

Attribute Type		Attribute Value	
Use	1	Personal name	1
		Title	4
		ISBN	7
		Date	30
		Author	1003
		Any	1016
		Publisher	1018
Relation	2	less than	1
		equal	3
		greater or equal	4
		greater than	5
		not equal	6
		AlwaysMatches	103
Position	3	first in field	1
		first in subfield	2
		any position in field	3
Structure	4	phrase	1
		word	2
		key	3
		year	4
		date (normalized)	5
		word list	6
		name (normalized)	101
		structure	103
		free-form-text	105
		local number	107
		string	108
Truncation	5	right Truncation	1
		left truncation	2
		left and right	3
		do not truncate	100
		process # in search term	101
		regExpr-1	102
Completeness	6	incomplete subfield	1
		complete subfield	2
		complete field	3

2.3 Z39.50 検索システム

2.3.1 システムの概要

本研究では日本語書誌データベースを検索するために Z39.50 に準拠したサーバとクライアントを UNIX ワークステーション Sun Enterprise 3000 上で開発した。サーバは Japan/MARC をデータとして持ち、Z39.50 に基づいたクライアントからの検索要求に応える。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が可能であり、Z39.50 に準拠した他のサーバにもアクセスできる。Z39.50 のクライアントとサーバは APDU(Application Protocol Data Unit) と呼ばれる情報フレームをそれぞれのプロトコルモジュール間で通信することで検索セッションを維持する [36]。

サーバはコントロールモジュール、プロトコルエンジン、検索エンジン、データベースから構成され、データベースはさらにインデックス部と Japan/MARC レコード部から構成される。要求は全てクライアントから APDU 形式で出され、サーバは受け取った APDU 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要求であれば検索エンジン呼び出す。検索エンジンはデータベースを検索し、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検索結果を履歴集合として保存し、ヒット件数をプロトコルエンジンに渡す。プロトコルエンジンはヒット件数をクライアントに返す。また、コントロールモジュールへのクライアントからの要求が返戻要求の場合、履歴集合として保存してあった検索結果を SUTRS(Simple Unstructured Text Record Syntax) 形式でクライアントに返戻する。

本システムは Index Data 社が開発した Z39.50 システム構築用ツールキット YAZ(Yet Another Z39.50 Toolkit)[37] 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。また、サーバの検索エンジンには全文検索エンジン Namazu[38][39] を用いた。クライアントのプロトコルエンジンは YAZ を用いて開発し、GUI 部分は Tcl/Tk を用いて開発した。

2.3.2 日本語書誌データの処理

本システムは日本語書誌データ Japan/MARC の検索を可能にするため、検索エンジン及びデータベースで日本語処理を行うとともに、クライアントで日本語入力および表示が行えるようにした。

この節では、Japan/MARC データを全文検索エンジン Namazu で利用できるようにするための切り出し処理およびインデックス処理について述べる。なお全文検索エンジン Namazu は、フリーで利用可能な WWW 上の検索エンジンとして広く使われており、本システムでは、これを Japan/MARC のタグフィールドに対応するように改造し、Z39.50 サーバで利用できるようにした。

本システムは 1983 年 ~ 1998 年 5 月までの配布分 1,019,696 レコードをデータとして持つ。国立国会図書館から配布された Japan/MARC レコードのデータはあらかじめ安斎作成のデコーダ

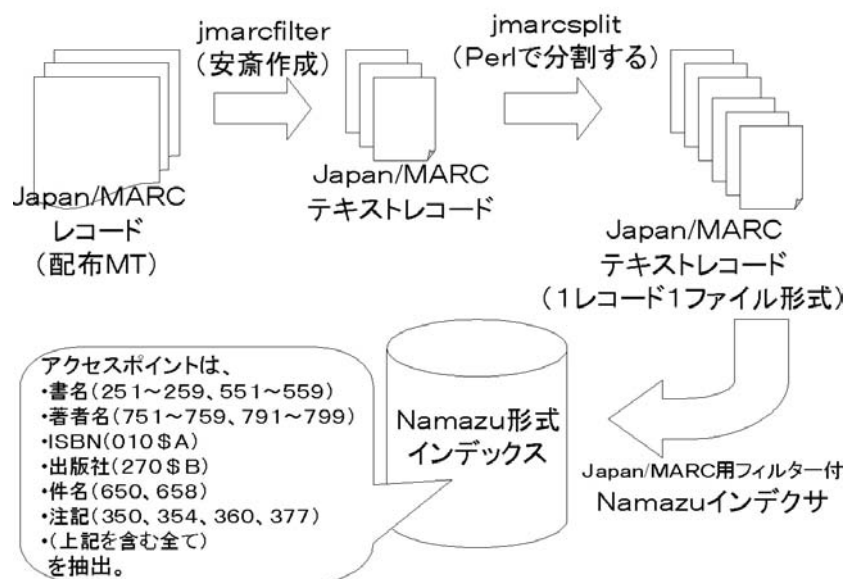


図 2.5 Japan/MARC の変換

[40] でプレインテキストにしたものをもとに 1 レコード・1 ファイル形式に変換する。このレコードファイルから Japan/MARC のタグフィールドの情報 [41][42] をもとに、ISBN、書名、著者名、件名、出版者、注記ごとのインデックスと全アクセスポイントを対象にしたインデックスを作成した(図 2.5 参照)。この際、日本語のテキストデータは日本語形態素解析ツール・茶筌 [43] でわかち書きを行なったものを Namazu のインデックスに登録しておく。インデクシングの所要時間は 14 時間 25 分 16 秒で、茶筌で切り出したキーワード数は 4,414,062 語であった。

Z39.50 以前の検索システムではシステム毎に検索質問の構造が異なり、複数のデータベースを利用する際の障害となっていた。Z39.50 ではこの問題に対して、検索質問の包括的なスキーマを定義し、実装の際にこのスキーマ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 つの数字の組み合わせで表される Bib-1 アトリビュートセットで定義されている。例えば、Bib-1 を使えば、書名・著者名・前方一致・後方一致・フレーズなどの検索質問の特性を特定できる。本システムではこのアトリビュートセット Bib-1 を用いた。Bib-1 と Japan/MARC でのタグ・フィールドとの対応関係を表 2.2 に示す。

2.3.3 Z39.50-ULIS3 サーバ

Z39.50-ULIS3 サーバは接続機能、検索機能、返戻機能、終了機能の基本機能を持つ。接続機能は複数の Z39.50 クライアントによる接続要求を受け付け、検索・返戻時に用いる推奨メッセージ

表 2.2 Bib-1 と Japan/MARC の対応表

項目	Bib-1(USE Attribute)	Japan/MARC タグ	Japan/MARC 説明
ISBN	ISBN(1=7)	010\$A	: 国際標準図書番号 (ISBN)
書名	Title(1=4)	251 ~ 259 \$A \$B \$D 551 ~ 559 \$A \$X \$D	: 記述フィールド : 書名 : 副書名 : 巻次等 : 書名アクセスポイント : カタカナ形 : ローマ字形 : 巻次の読み
著者名	Author(1=1003)	751 ~ 759 \$A \$X \$B 791 ~ 799 \$A \$X \$B	: 著者名アクセスポイント : カタカナ形 : ローマ字形 : 漢字形 : 多巻ものの各巻著者標目 : カタカナ形 : ローマ字形 : 漢字形
件名	Subject Headings (1=21)	650 658	: 個人名件名標目 : 一般件名標目
出版社	Publisher (1=1018)	270\$B	: 出版者、頒布者等
注記	Note (1=63)	350 354 360 377	: 一般注記 : 原タイトル注記 : 装丁と定価に関する事項 : 内容注記
全て	Any (1=1016)	数字 3 桁\$.	: タグ・フィールドを除いた全て

サイズやサーバが提供している機能、Z39.50 のバージョン情報などのパラメータの折衝を行う機能である。

検索はコントロールモジュールがプロトコルエンジンから検索要求を受けとり、コントロールモジュールが逆ポーランド記法の検索式を検索エンジンに渡すことによって行われる。検索エンジンはインデックスを検索して、コントロールモジュールに検索結果を返す。コントロールモジュール

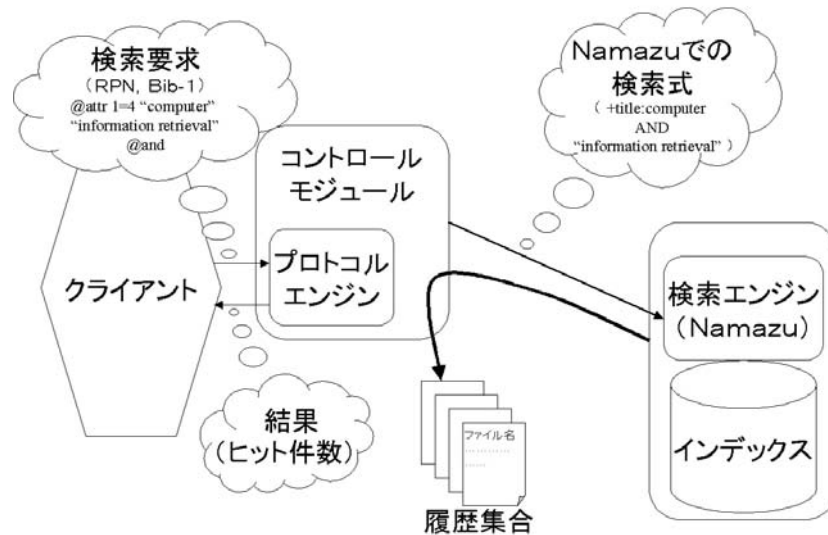


図 2.6 検索処理

ルはヒットしたレコードのファイル名を履歴集合ファイルに出力した後、ヒット件数をプロトコルエンジンを通してクライアントに返す（図 2.6 参照）。サーバは検索式として Type1-Query と Type101-Query を受けつける。検索質問は AND, OR, AND-NOT の論理演算、Bib-1 のアクセスポイント指定、検索語の属性指定、履歴集合指定が可能である。履歴検索は履歴集合同士や、検索語と履歴集合の論理演算を行える。履歴集合同士の検索は履歴ファイルのみで行うため短時間で処理できる。

返戻はまず履歴集合名、レコードシンタックス、返戻レコード開始番号、レコード件数を指定し、指定された履歴集合から実際の Japan/MARC レコードファイルを抜き出し、Japan/MARC レコードファイルを SUTRS 形式で返すことによって行われる（図 2.7 参照）。クライアントへ返戻するレコードのサイズは、接続要求時の推奨メッセージサイズのパラメータ指定による。

終了機能はクライアントとの接続を切り、そのセッションでの履歴集合ファイルの削除を行なう機能である。

2.3.4 Z39.50-ULIS3 クライアント

クライアントの役割は GUI を通したユーザからの接続要求、検索要求、返戻要求を解析・変換し、サーバに対して出すことである。それぞれの要求はプロトコルエンジンによって APDU に変換されてサーバに渡され、サーバから返された APDU は出した要求に応じてクライアント内で処

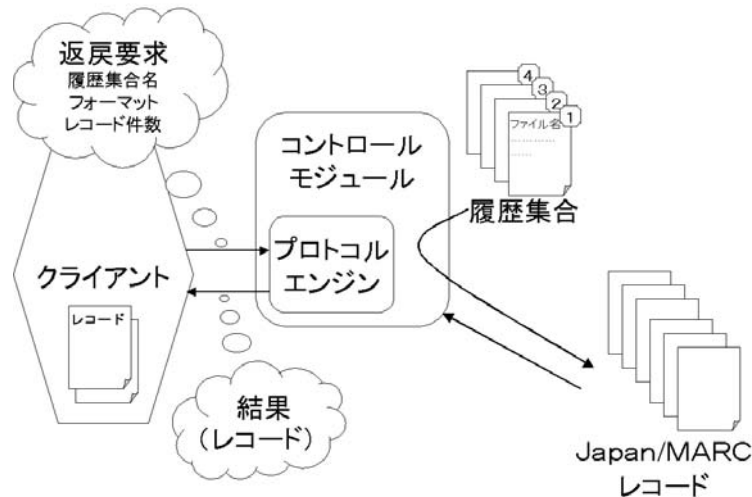


図 2.7 返戻処理

理される。ユーザは本検索クライアントを用いて著者や書名などのアクセスポイントを指定したり、前方一致や後方一致、履歴を利用して検索できる。また、レコード数、履歴集合、詳細度合を指定して検索結果を見ることができる。

図 2.8 は接続ウインドウの画面である。画面中央には OCLC First Search や University of California など約 200 のサーバのリストが表示されており、ユーザが接続したいサーバを選択して connect ボタンをクリックすると、選択したサーバの IP アドレス、ポート番号、データベース名が画面上部のウインドウ内に入力され、接続が開始される。リストにないサーバに接続したい場合は、キーボードから直接サーバ情報をウインドウに入力できる。また、サーバ情報を表す URL を指定ファイルに登録することによりリストにサーバ名を追加できる。さらに、現在は Index Data 社が作成しているサーバリスト [20] を定期的に参照することによりサーバリストを自動的に更新しており、エンドユーザは常に最新の情報に基づいてサーバを選択できる。

図 2.9 は検索ウインドウ画面である。ユーザは検索式の入力、検索ボタンのクリック、検索結果の確認、結果集合の指定、返戻ボタンのクリック、返戻結果の確認という一連の操作を検索ウインドウ上で行う。その際、入力フォームやボタンがユーザの操作に連動して色が変わり、ユーザが次にどの操作をすべきかがわかるようになっている。

ユーザが検索式を入力し、search ボタンをクリックすると、検索結果集合番号、検索式、ヒット件数が表示される。新しく検索式を入力したい場合には clear ボタンをクリックすればよい。検索結果の確認後、検索結果の内容を見たい場合は、表示させたい検索集合番号をクリックして show

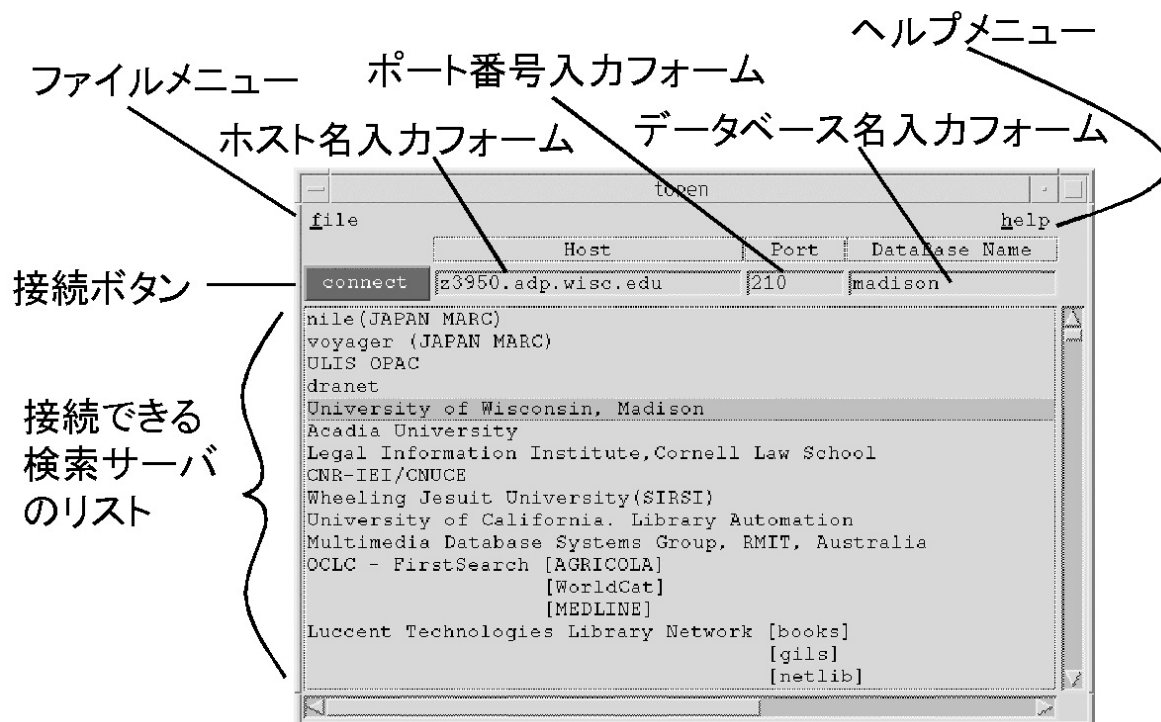


図 2.8 接続ウインドウの機能

ボタンをクリックすると指定した検索集合のレコードが表示される。検索集合を指定せずに show ボタンをクリックすると、直前の検索結果集合のレコードが表示される。また、キーボードから検索結果集合番号、開始レコード番号、終了レコード番号の直接指定もできる。

検索式の書式は検索コマンドの標準規格 (ANSI Z39.58-1992) である CCL(Common Command Language) を採用した。これはもともと YAZ が CCL から逆ポーランド記法の検索式に変換する [44] 機能を持っているためである。返戻結果の表示は Brief、Full ボタンをクリックすることで、簡易表示 (Brief)・詳細表示 (Full) を選択できる。また、返戻形式メニューから USMARC や SUTRS などのレコード形式を選択できる。

ファイルメニューの「New Window」を選択すると、現在接続しているサーバに接続したまま、新しい接続ウインドウを起動させ、他のサーバに接続できる。接続メニューの「Disconnect」を選択すると、現在接続しているサーバと接続を切り、新しい接続ウインドウが起動し、他のサーバに接続できる。ファイルメニューの「Exit」を選択するとクライアントを終了する。ヘルプメニュー

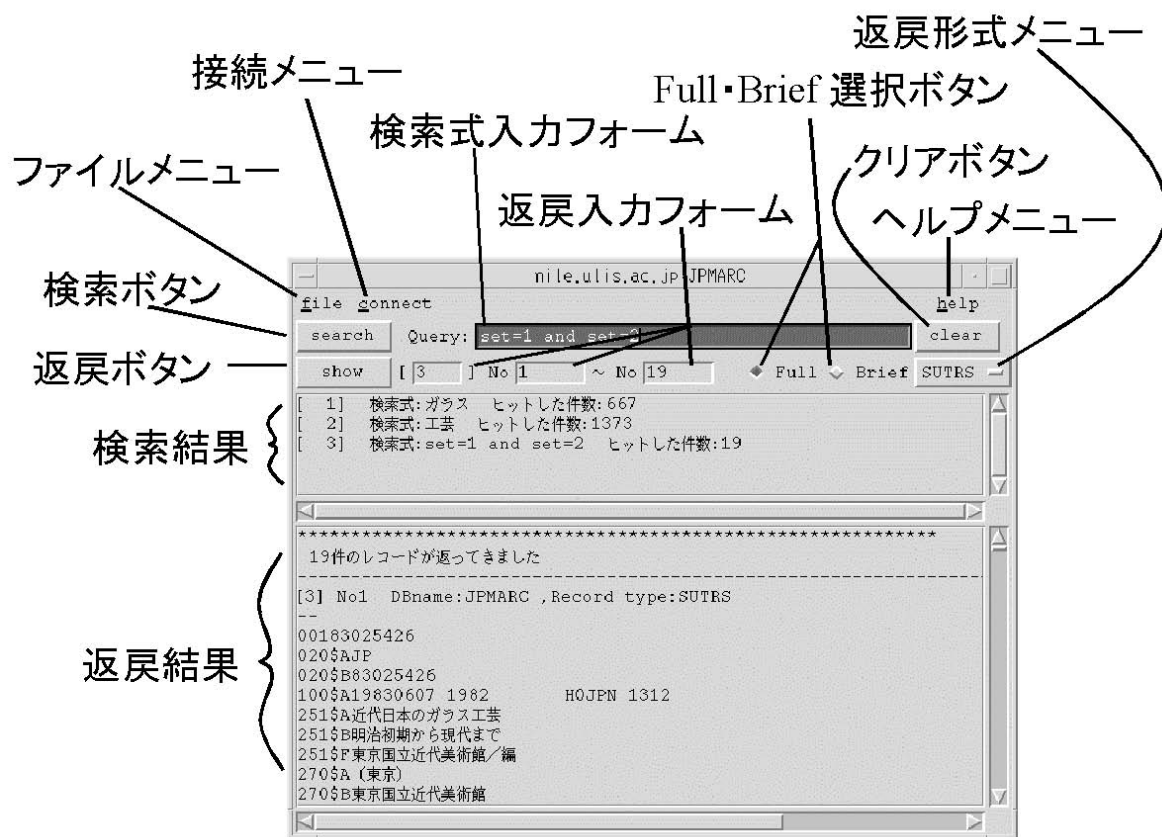


図 2.9 検索ウィンドウの機能

を選択すると、検索式の例が表示される。

2.3.5 システムの実行例

これまで述べてきたように、Z39.50-ULIS3 システムは Z39.50-ULIS3 クライアントと、Z39.50-ULIS3 サーバの 2 つの独立したアプリケーションからなっている。この節では、Z39.50-ULIS3 クライアントを用いて Z39.50-ULIS3 サーバを含めたいくつかの Z39.50 サーバを検索した実行例をあげることで、Z39.50-ULIS3 サーバや Z39.50-ULIS3 クライアントが日本語を扱えることや、Z39.50-ULIS3 クライアントが Z39.50-ULIS3 サーバだけでなく他の Z39.50 に基づいたサーバを検索できることを示す。

まず、Z39.50-ULIS3 サーバを検索した例を図 2.10 に示す。検索結果表示欄を元に検索を順に追うと、



図 2.10 検索ウインドウ : Z39.50-ULIS3 サーバ

1. 「ガラス」を検索 ヒットした件数が 667 件
2. 「工芸」を検索 ヒットした件数が 1373 件
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 19 件

という検索要求が出されていることがわかる。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。

次に、Z39.50-ULIS3 以外の日本語を用いたサーバを検索している例を図 2.11 に示す。図 2.11 は図書館情報大学デジタル図書館システム Z39.50 サーバを検索しているところである。検索結果表示欄を元に検索を順に追うと、

1. 「統計」を検索 ヒットした件数が 1425 件
2. 「書名に統計」を検索 ヒットした件数が 1154 件
3. 「書名に統計」と「SPSS」の AND 検索 ヒットした件数が 5 件

という検索要求が出されていることがわかる。2 番目と 3 番目の検索式の“ti=”はアクセスポイントに書名を使うことを表している。3 番目の検索式の“and”は「書名に統計」と「SPSS」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。

最後に海外のサーバを検索しているところを図 2.12 に示す。図 2.129 は University of Wisconsin,



図 2.11 検索ウインドウ：図書館情報大学デジタル図書館システム Z39.50 サーバ

Madison のサーバを検索しているところである。検索結果表示欄をもとに検索を順に追うと、

1. 「書名に library」を検索 ヒットした件数が 5948 件
2. 「著者名に Lancaster」を検索 ヒットした件数が 498 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

という検索要求が出されていることがわかる。1 番目の検索式の“ti=”はアクセスポイントに書名を使うことを表しており、2 番目の検索式の“au=”はアクセスポイントに著者名を使うことを表している。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。

2.3.6 開発システムの変遷

今回開発した Z39.50-ULIS3 システムと以前の Z39.50-ULIS1 システム [45] および Z39.50-ULIS2 システム [46] との相違点を表 2.3 にまとめた。

Z39.50-ULIS1 は本学で開発した最初の Z39.50 システムであり、Z39.50 システムを構築するための基本技術が得られた。しかしながら、この時点ではインターネット上に日本語書誌データを提供することが主要な目的であったため、利用者に馴染みのある WWW ブラウザをクライアントと

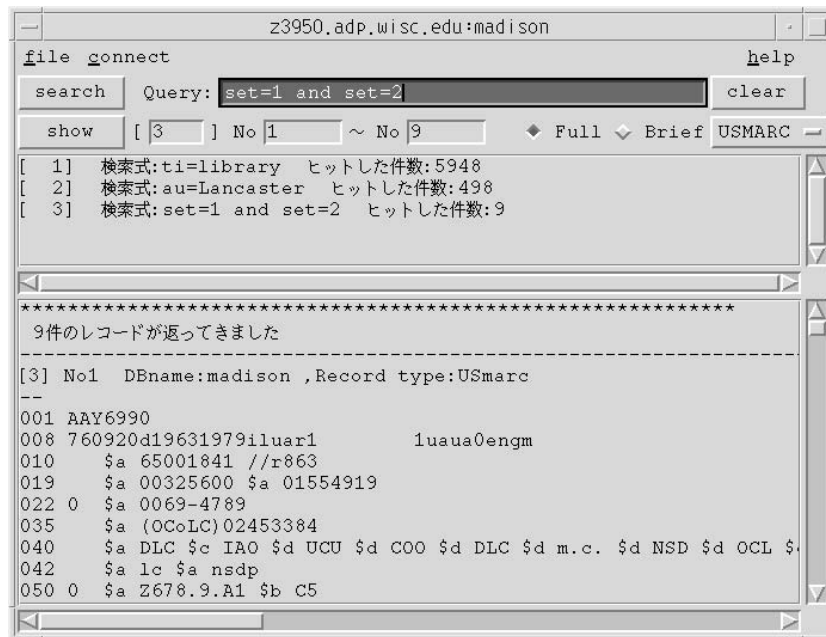


図 2.12 検索ウインドウ：University of Wisconsin, Madison

し、CGI を介して検索するシステムとして構築した。そのため、純粋のクライアントとサーバの両方を含む完全な Z39.50 システムにするには至らなかった。

Z39.50-ULIS2 は Z39.50-ULIS1 のシステム構築の基本技術を継承し、Z39.50-ULIS1 では実現しなかった Z39.50 の仕様に完全にそったシステム構築を目指し、純粋な Z39.50 クライアントを作成するとともにサーバも新たに構築し直すことにした。しかしながら、実装面でいくつかの不具合があり、検索システムとして見た場合、不満が残るシステムであった。

Z39.50-ULIS3 は Z39.50-ULIS2 と同じクライアント・サーバ型のシステム構成であるが、Z39.50-ULIS1 と Z39.50-ULIS2 で得た技術を統合し、サーバの実装面で多くの改良を加え、設計と実装の両面においてより充実したシステムとなった。

Z39.50-ULIS3 サーバでの主な改良点は以下の 2 点である。

1. 大規模システムの構築（データ量の増大）

本システム（Z39.50-ULIS3）では約 100 万件のデータを Z39.50 で提供する基盤を構築するために検索レスポンスの面での大幅な改善を求めた。そのために、ULIS2 のシステムの見直しを迫られ、Namazu を中心とする全文検索エンジンの導入に至った。検索実験も約 100 万件のデータに対して行い、支障のないレスポンスタイムで検索できた。

2. Namazu の改造による資源共有

フリーで入手可能な Namazu に対して Japan/MARC に対応させるための改造を行なったこ

表 2.3 開発システムの相違点

	Z39.50-ULIS1 (1996 年 11 月)	Z39.50-ULIS2 (1998 年 2 月)	Z39.50-ULIS3 (1999 年 5 月)
構成	サーバ WWW ブラウザから検索 (CGI 接続プログラム)	サーバ クライアント	サーバ クライアント
開発環境	Sun SPARC Station10 YAZ-1.2 DBMS: ADABAS C 言語	IBM RS6000 YAZ-1.3 DBMS: Oracle7 C 言語, SQL	Sun Enterprise 3000 YAZ-1.4pl2 検索エンジン: Namazu C 言語, Perl
データ	Japan/MARC 1 年分 (73,549 件) ULIS-OPAC (85,651 件)	Japan/MARC (1,601 件)	Japan/MARC 15 年 5 ヶ月分 (1,019,696 件)
検索機能	論理演算可 フィールド指定可 (書名、著者、ISBN、 件名、NDC、 NDLC, JP-No) 履歴検索可	論理演算一部可 フィールド指定可 (書名、著者、全指定) 履歴検索可 (履歴集合同士のみ)	論理演算可 フィールド指定可 (書名、著者、ISBN、 件名、出版者、 注記、全指定) 履歴検索可
レコード	ADABAS のレコード	Oracle のテーブル	1 レコード 1 ファイル
インデックス語 の切りだし	MHSA[47]	なし (SQL の LIKE)	茶釜

とにより、WWW 上で広く利用されている Namazu の検索システムの資源（インデックス）をそのまま Z39.50 で提供できる。

2.4 考察

以上、本研究で開発した Z39.50 検索システムのクライアントとサーバの構成と機能について述べてきた。ここでは、本システムの意義と Z39.50 検索システムの将来について考察する。

2.4.1 システムの意義

本システムの意義は実験システムという観点から、システム構築をとおして様々な試みができるという点である。Z39.50 はクライアントとサーバ間のやりとりのみを規定しているので、クライアントとサーバ自身は自由に構築できる。しかし、自由に構築できる反面、自ら構築しなければならず、Z39.50 システムを必要としているが開発に不慣れな場合はかなりの負担となる。そこで、Z39.50 システムはどういう性格のシステムなのか、どういうシステム構成にすればよいのか、どのくらい手間がかかるのか、必要な機能は何で、これからどういう機能が必要になるかなどを実験をとおして知る必要がある。

システム構築はシステムの性格、データの性質、開発の利便性などを考慮に入れて行った。Z39.50 検索システムは自由度が大きい部分と制限の強い部分を同時に合わせ持つシステムであり、機能的には相互接続性を保証すること、セッションを維持することが重要である。その他、検索レスポンス、日本語処理、検索式入力形式、履歴検索、横断検索なども考慮する必要がある。本システムで扱う書誌のデータは、量が多いこと、処理単位が基本的にレコードであること、ひとつのレコードが構造を持つこと、検索の際は複数のレコードにまたがってアクセスすること、データ更新のほとんどが追加のみであることなどが性質としてあげられる。また、開発の利便性については、どこまで既存の道具を使い、どこから独自にプログラムを書くのか、開発後の移植性などを考慮しなければならない。

以上のようなことを踏まえ、サーバについては、ファイル編成法やデータアクセス手法について実験を繰り返した結果、レコードファイルとインデックスを完全に分離し、レコードファイルを直接処理するのではなく、レコードファイルに対してアクセスポイントごとにインデックスをつけ、インデックスファイルを直接の処理対象にするという手法が最も効率的であることがわかった。関係データベースで直接書誌データを処理することも試してみたが、従来の書誌データ検索システムで使われてきたオーソドックスな手法が最も有効であることが明らかになった。

Z39.50 の実装については、基本的な機能については実現したが、細かい部分については今後の課題として残っている。サーバに関していうならば、例えば、Bib-1 のアトリビュートタイプの内、アクセスポイント指定 (Use) 以外の Relation、Position、Structure、Truncation、Completeness の指定には対応しておらず、ElementSetName(詳細・簡略)の指定にも対応していない。Sort や Scan などの拡張機能についても未対応である。しかし、これらの機能は書誌データ検索としては必須の

機能ではないと判断し、今後の課題とした。

文字コードについては、サーバとクライアントの内部では EUC で処理し、Japan/MARC は ISO-2022-JP で保存してある。そのため、サーバからクライアントへのレコード返戻時には ISO-2022-JP で転送する。ただし、クライアントは ISO-2022-JP だけでなく Shift-JIS と EUC のデータが来ても EUC に自動変換して文字化けせずに表示できるようにしてある。サーバ側もクライアントからの入力文字コードを内部文字コードである EUC に変換して処理している。このため、日本語文字については言語折衝 [48] がなくても、日本語データを持った Z39.50 サーバを検索できる。将来的には言語折衝の部分も実装したいと考えている。

クライアントのその他の機能実験としては、Sun のワークステーションのほか、IBM と Hewlett-Packard のワークステーションにも移植し、マルチプラットフォーム化を行った。また、サーバの自動登録機能についても開発した。

今後、サーバ情報を利用者同士や他システム間で交換することを想定して、接続時のサーバ情報は、`z39.50s://lib.ulis.ac.jp:210/opac` (図書館情報大学デジタル図書館) のように標準化された URL 形式 [49] で記述した。

このようにシステム構築を通して様々な試みを行っており、今後も引き続き実験を行う予定である。

2.4.2 Z39.50 検索システムの将来

本節では、Z39.50 検索システムの将来について論じ、今後の実験の方向性を探る。

検索システムに Z39.50 プロトコルを採用することのメリットは、使い慣れたクライアントで複数のサーバに接続できること、横断検索が可能になること、セッションを維持した履歴検索ができるという 3 点にまとめられる。

検索システムを相互に接続し、書誌データを交換しようという 1979 年からの計画は、1992 年の Z39.50 Version2 の規格が制定されたのを機に実装が進み、これから順調に発展していくように見えた。ところが、1994 年からの World Wide Web(WWW) の爆発的な普及により、Z39.50 の存在意義を疑う意見が出てきた。しかし、WWW はもともとは研究者間の研究情報を共有するという目的で開発されたブラウジング指向のシステムであり、WWW ブラウザ上のハイパーリンクをたどることによって、関連情報を次から次に見ることができる。ハイパーリンクは HTML と呼ばれるマークアップ言語によって記述し、HTML ファイルは WWW 用の転送プロトコル HTTP によって転送される。

WWW が単なるブラウジングシステムに留まっていたならば、Z39.50 とは競合せず、比較されることはなかった。ところが、ブラウジングだけでは求める情報にたどり着くのが困難になり、検索エンジンがブラウジングシステムとは別に実装された。また、検索エンジンの他にもユーザとインタラクティブに動作するシステムが数多く提供された。こうなると、WWW はもはや単体の

ブラウジングシステムというより、複数のシステムが提供される総合環境と位置づけられる。

その結果、本来 Z39.50 に適していた図書館 OPAC も WWW ベースで提供されることとなり、利用者は使い慣れた WWW ブラウザから OPAC を利用できるようになった。しかも、それぞれの OPAC のユーザインタフェースは似通っており、使い勝手が阻害されるということもない。そのうえ、九州大学の WWW ページ [50] では複数の OPAC に対して横断検索もできるようになっている。

このように、Z39.50 のメリットは大部分 WWW で実現されてしまった。特に日本では、Z39.50 の普及より WWW の普及のほうが先であったため、より WWW の優位性が目立つ結果となった。しかしながら、これで WWW があれば Z39.50 が必要ないということにはならない。上田も述べているように [16]、現実的観点からみると WWW のほうが有利であるが、視点を変えることによって Z39.50 の重要性が浮かび上がってくる。

筆者が考える Z39.50 の重要性は次の 3 つである。

1. 既に数多く立ちあがっている海外サーバの利用および国内サーバの海外への提供
2. 複雑な検索要求には履歴検索が有効
3. 構造情報を含めたデータ交換
 - (a) WWW とは違った検索に特化したネットワークの可能性
 - (b) 電子図書館システムにおける構造情報通信のための基盤技術

まず 1 点目についてであるが、欧米では WWW が普及する前に Z39.50 の実装が進んだため、既に数多くの Z39.50 サーバが立ち上がっている。これらの資源を利用することはたいへん有用であり、前述のとおり、日本でも海外サーバに対してゲートウェイサービスが行われている。今後は海外のサーバを利用するだけでなく、日本でもサーバを立ちあげて世界に日本のデータを提供することが重要である。

次に 2 点目であるが、WWW の検索エンジンにより数多くの人が自ら手軽に検索できるようになり、このような検索エンジンがあれば検索はこれで全て間に合うという考えが増えてきた。しかし、検索要求や検索対象は質的にも量的にも多様性があり、WWW コンテンツ、図書、雑誌記事、ファクト、数値データの検索を同列に扱うべきではない。確かに、WWW コンテンツの検索は検索エンジンで用が済む場合が多いが、これはもともとの検索要求が比較的緩いことと、ブラウジング機能も併用しているので、検索がうまくいっているかのように見えるのだと思われる。この場合でも、適合度順つきの検索結果と実際の検索要求が一致しないなど WWW 検索は多くの問題を抱えている。そして、より高い検索精度が要求される雑誌記事を対象にした場合はさらに深刻で、複雑な検索式を用いた繰り返し検索が必要になってくる。そうなった場合、履歴検索は必須の機能と言えるだろう。

最後に 3 点目であるが、WWW で使われている通信プロトコル HTTP と Z39.50 との本質的違いは、HTTP が基本的には HTML ファイルの単純転送であるのに対して、Z39.50 はデータ構造や

処理方式を含んだ情報通信であるという点である。現在、WWW コンテンツとは異なる形式の多量のデータを管理するため、独自の検索エンジンを持つ人が増えつつある。これまでは特定の検索エンジンを利用していたのが、独自の検索エンジン込みでそれぞれがデータを提供できるようになる可能性がある。そうすると、お互いのデータを交換し、再利用したいという要求がでてきてデータ構造を理解する共通のプロトコルが必要になる。この共通のプロトコルに Z39.50 を使えば、クライアント機能を内部に持ったサーバ同士が、Z39.50 を介して直接データを交換できる。こうなると、Z39.50 は検索プロトコルというよりデータ交換プロトコルであると言えるだろう。その結果、Z39.50 は構造データが対象の検索に特化したシステム、WWW はコンテンツの提供を主目的にしたシステムというように、性格が異なるシステムとして共存することが可能であり、それぞれの性格を活かした関係も可能になる。

その他、構造を持ったデータ交換が可能ということで、電子図書館などのシステムで内部通信のプロトコルに使える。例えば、密結合が要求されるサブシステム間のデータ交換などシステム間の共通データ通信プロトコルとして利用することは大いに価値があると思われる。通信の部分は仕様が決まっている Z39.50 を使えばプログラム開発が楽にでき、その分、他の部分の開発に集中できる。

現在、Z39.50-1995(Version 3) の ISO 版である ISO 23950 が JIS 規格としても制定されており [7]、日本でも今後ますます実装が進んでいくだろうと思われる。

次に、Z39.50 におけるプロファイルについて述べる。Z39.50 仕様を管理している ZIG の Web サイト [51] によれば、「プロファイルは、特定の

- アプリケーション: GILS・WAIS など
- 機能: 著者/タイトル/主題の検索など
- コミュニティ: 博物館・化学・音楽など
- 環境: インターネット・北米・ヨーロッパなど

に対応するため、どの標準のどの部分を利用するかを指定する」ものであり、「利用方法の指定とは、Z39.50 標準でデータベース毎に実装が任されている範囲の、オプション、サブセット、パラメータ値を適宜選択することを意味する」と定義されている。Z39.50 の仕様には様々な機能についての取り決めがあり、実際に提供されている Z39.50 システムが持つ機能には幅がある。例えば、接続や終了、検索、返戻などの Z39.50 システムにおける基本機能のようにほとんどのシステムが実装する機能の他にも、各システムにより実装されているかどうかで別れる様々な拡張機能がある。さらに、実装されている機能であっても、そのデータベースが使用している、アクセスポイントや検索結果のデータ形式などに差異がある場合も多い。そのため、Z39.50 を特定の用途で使用する際に、そのコミュニティにおいて Z39.50 システム運用上の合意事項を取り決める必要が出てきたため、プロファイルを作成してこの合意事項を指定する動きが出てきた。Z39.50 仕様書 [7] にはその出版時点で、Open Systems Environment Implementors Workshop(OIW) に承認されてい

る4つのプロファイルが掲載されている。その4つとは、GILS, WAIS, ATS-1, Using Z39.50-1992 Directly over TCPである。さらにその後も、ATS-1などの取り組みを引き継いだ形で、図書館間アプリケーションを構築するのに必要な相互運用性の確保のための Bath プロファイル [52] といったプロファイルの標準化が進んでいる。

本研究では、特定のプロファイルに従ったクライアント・サーバの開発は行なっておらず、多くの Z39.50 システムの基本的な利用を可能にするために必要だと筆者が判断した機能を選択し構築した。Z39.50 の利点の一つに使い慣れたクライアントから複数のサーバに接続できることを挙げたが、このように様々なプロファイルが存在する場合、同じ Z39.50 で提供されているデータベースでも、使い勝手が大きく違ってしまったり、接続することができないなどの問題が出てくる。例えば、Using Z39.50-1992 Directly over TCP プロファイルのように異なるネットワークで提供するためのプロファイルを採用したシステムの場合は、本研究で開発したクライアントでは接続できない。これは、本研究で開発したクライアントが劣っているというわけではなく、そのプロファイルに対応していないというだけであり、プロファイルが、各コミュニティにおいての Z39.50 システム運用上の合意事項であるという性質上、全てのプロファイルに対応するシステムを構築することは現実的ではない。

そのため、Z39.50 システムを開発する場合、システムの用途に合わせて多くの Z39.50 システムの基本的な利用が可能なシステムにするのか、それとも特定のプロファイルに対応したシステムにするのか、それとも多くのプロファイルを切り変えて使えるシステムかなど、プロファイルにどう対応するのか判断が必要となる。また、どのプロファイルを採用しているか、対応しているプロファイルが何かなどの情報を、利用者にわかるようにしておく必要も出てくるだろう。

2.5 まとめ

本研究では検索ユーザインタフェースの統一という観点から、情報検索標準プロトコル Z39.50 に基づいた日本語書誌データ検索システムを構築した。Z39.50 を使うことにより、ひとつの検索クライアントだけで、世界中に存在する多くの Z39.50 サーバが検索でき、検索ユーザインタフェースの統一が可能なことを示した。また、日本語入力・日本語表示が可能な Z39.50 クライアントを開発することにより、Z39.50 において、日本語を扱った検索サービスが提供可能であることも示した。

欧米だけでなく日本でも、1999 年の東京工業大学や図書館情報大学の電子図書館システムなど、自館の OPAC データを Z39.50 でサービスするシステムが出てきた。本研究では、日本語入力・日本語表示という日本語を扱う際の最も基本的な機能が実現可能であることを示したが、これ以外にも、Z39.50 を使ったサービス提供のために必要な解決すべき問題は数多く残っている。例えば、海外サーバを含む他のサーバとの相互接続性、レコードシンタックスおよびセマンティックス、メタデータと既存目録データの交換方式、フルテキストなど書誌データ以外の処理機能、また、多言

語を扱う際の問題としては、文字コードが異なった場合の対処方法などもある。このような状況の中で、様々な実験を試せる実験システムという側面においても、本システムの意義は大きい。

第 3 章

Z39.50 データベース選択支援環境

3.1 はじめに

1970 年代から数多くのデータベースが登場したが、上田 [16] が指摘するように、「それぞれのデータベースシステムがそれぞれ異なり、同じデータベースを使う場合でもそれぞれのシステムのアクセス方法やコマンドやデータベースの構造を覚えなければならない」「高額な利用料やそれぞれの提供機関と契約を結ばなければならない」といった問題があり、一般の利用者が多くのデータベースを検索できる状況にはなかった。

しかし、情報検索を透過的に行うための標準プロトコルである Z39.50 が登場したことによりこの状況は変化した。Z39.50 は 1988 年に ANSI 標準となり、1992 年・1995 年 [14] と改訂が加えられ、欧米の大学図書館を中心に広まった。ここ数年、日本でも電子図書館システムでの採用を契機として急速に普及し始め、利用者は異なるシステムを同じユーザインタフェースで検索できるようになった。

その上、コンピュータ技術の進歩によってコンピュータが安価で手に入れられるようになり、様々なデータベースの検索サービスも無料で提供されるようになってきた。また、インターネットの普及により、どこにいてもデータベースにアクセスできるようになり、分散型データベースが多く提供されるようになった。この結果、利用者がシステムや場所を意識せずに検索できるデータベースが急激に増加している。

以前であれば、データベースのほとんどは DIALOG のような商用で大規模な集中型の検索システムだったため、データベースの利用者は、ほとんどがサーチャーや図書館員などの検索のスペシャリストであった。彼らはデータベースベンダなどが開く講習会や、それらのデータベースの特徴や使い方について書かれた文献（例えば化学分野では文献 [53][54] 薬学分野では文献 [55]）から得た知識や自身の経験を元に適切なデータベースを選択できた。しかし、様々な分散型の検索システムが存在し、一般利用者が多くのデータベースを検索できるようになった現在では、検索に熟練していない利用者が適切なデータベースを選択しなければならず、必要なデータベースを選び

損なったり、あるいは逆に、不要なデータベースを選択、検索することにより、システムやネットワークにいたずらに負担をかける [56] などの問題がでてきた。

分散型の検索システムは、検索システムをそれぞれの機関が作成しているために、データベースに関する情報をまとめて得ることが難しく、また情報の変化が速い。例えば、医学・薬学に関する単行本 [57] のようなインターネット上のデータベースの検索の方法を解説した文献もあるが、情報の速い変化に本の改訂で対応するのは困難である。

そこで、我々はこの問題を解決するために Z39.50 データベースの選択支援環境が必要になると考えた。

データベース選択支援を目指した既存のシステムには、Cornell 大学の Gateway システム [58][59] のようにデータベースを分野ごとに分類し、WWW-Z39.50 ゲートウェイを利用して Z39.50 データベースを提供しているサイトがある。このようなサイトを利用すれば、ユーザは組織化されたデータベースの中から適切なデータベースを選んで検索できる。しかし、WWW-Z39.50 ゲートウェイを用いたシステムの場合、ユーザ個々人の方針による組織化やデータベースの追加ができない。

データベースを選択するには、ユーザの要求にあったデータベース情報の収集や組織化が必要だが、全てのユーザに適した収集や組織化を行うことは難しい。そのため、ユーザ毎にカスタマイズできる必要がある。

また、ネイティブな Z39.50 クライアントでは、ユーザによるカスタマイズは可能だが、データベース選択のための情報を共有することは難しい。個々のクライアントに登録されたデータベース情報は、ユーザ自身しか利用できないため、個々のユーザがデータベース情報をクライアントに登録し、組織化しなければならないからである。

そこで、我々は Z39.50 書誌検索システム [10][34][60] を構築した経験を基に、本研究ではデータベース選択支援環境を提案し、実装した。この環境では、ユーザのデータベース選択を支援するため、Z39.50 データベースに関する情報を WWW 上で共有して利用できるだけでなく、利用者が自由にデータベースを組織化し利用しかつ、利用者同士でお互いに参照・利用できる。

3.2 データベース選択支援環境

本研究で提案するデータベース選択支援環境では、データベースに関する情報を共有でき、共有された情報の中からデータベースを選択してすぐに検索に移れる（図 3.1）。この環境はいわば WWW と Z39.50 の世界が透過的に結合されているといえる。そこで我々は、WWW と Z39.50 が結合されていることから単なるシステムではなく、環境と呼ぶ。

WWW は情報を共有すること、すなわち、個人や組織が作成したコンテンツをお互いに利用できる点に優れ、一方、Z39.50 はさまざまな機関や組織が提供しているデータベースを共通の手順で検索できるなど、透過的な情報検索を行える点で優れている。そこで本環境では、情報共有に WWW を用い、情報検索に Z39.50 を用いた。

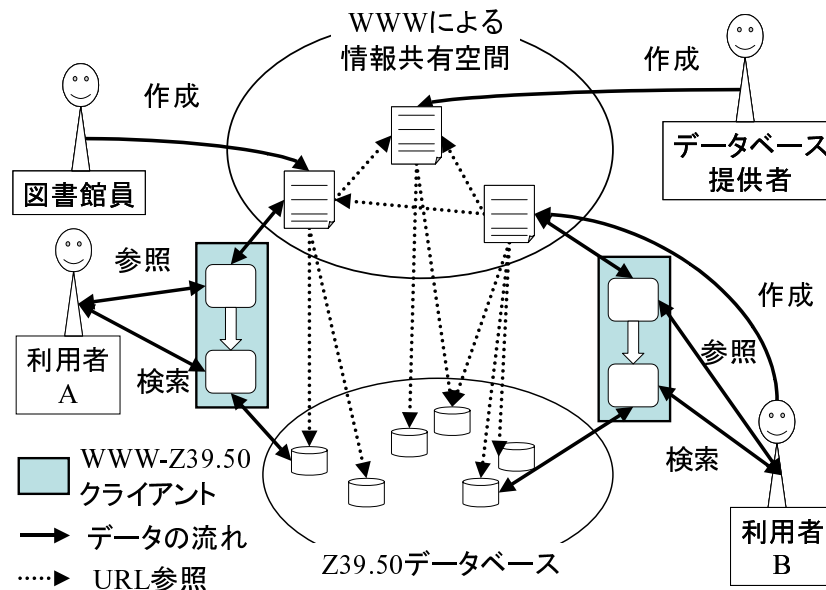


図 3.1 データベース選択支援環境の全体像

本研究では、WWW で共有された情報を閲覧し、ハイパーリンクをたどるクリック一つで Z39.50 データベースの検索に移れる WWW-Z39.50 クライアントを開発することにより、このデータベース選択支援環境を実現した。

本環境を説明する前に、WWW での情報共有について説明する。WWW ではリンクを辿ることによりさまざまなページにたどり着くことができる。ウェブページには、Yahoo![61] のように網羅的にウェブページを収集し分類してあるものや、サブジェクトゲートウェイのように、ある特定の分野を収集してあるもの、個人のリンク集のように良く使うものを収集してあるものなどがある。これらのウェブページは多くのウェブページの URL を集めたものであり、集めた URL を後から使いやすいように、検索できるようにしたり、並び替えたり、分類したり、その URL が示すウェブページの説明を付けたりしている。これらのウェブページは、作成者が使えることはもちろん、誰でも利用することができる。このように、WWW では利用者同士の情報共有が可能である。

本環境では、全ての利用者はウェブページのリンクをたどるだけで Z39.50 データベースが検索できるため、データベースに関する情報共有も可能となる。そこで、単なるウェブページだけでなくデータベースの情報を含めた情報の共有が可能な空間を「WWW による情報共有空間」と呼ぶ。また、Z39.50 データベースへのリンクが収集されており、だれでもリンクをたどるだけでそのデータベースに検索できるウェブページを「データベースリンク集」と呼ぶ。

例えば、図 3.1 において利用者 A は WWW による情報共有空間上の図書館員の作成したデータベースリンク集をたどって適切なデータベースを選び、検索できる。また、図書館員の作成したり

リンク集からはデータベース提供者が作成したリンク集にもリンクが張られているため、利用者 A はそのリンク集からもデータベースを検索できる。このように、利用者は自分でデータベースのリンク集を作成しなくても、図書館員など他の人が作成したリンク集を利用してデータベースを検索できる。また、利用者 B のように、自分用のリンク集を作成して利用することもできるため、利用者が自分の用途に合わせてデータベースを組織化することも可能となる。

このように、筆者は、いままで独自に普及してきた WWW と Z39.50 の両者を本研究で開発した WWW-Z39.50 クライアントで結合することにより新たなデータベース選択支援環境が実現できると考えている。

3.3 WWW-Z39.50 クライアント

3.3.1 WWW-Z39.50 クライアントの概要

本研究で提案するデータベース選択支援環境は以下のシステムを含む（図 3.2）。

1. Z39.50 データベース接続情報を表す URL を扱う WWW ブラウザ
2. この WWW ブラウザから起動される Z39.50 クライアント

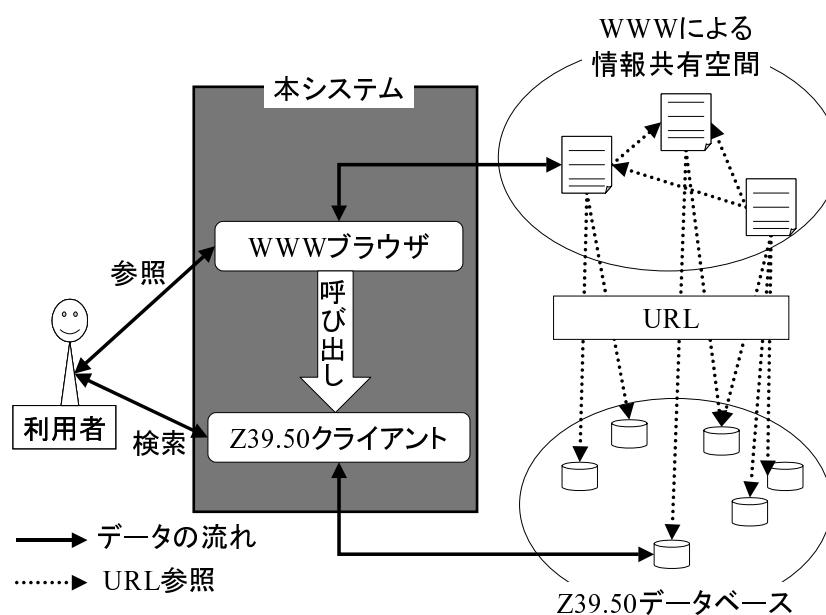


図 3.2 システムの全体像

URL とはインターネット上に存在する様々なリソースをあらわすためのアドレスであり、WWW で広く利用されているが、本環境ではデータベース接続情報を URL で記述することにより、WWW

```

<html>
<head>
<title>医学関連のデータベースリスト</title>
</head>
<body>
<h1>医学関連のデータベースリスト</h1>
  OCLC の無料アクセス可能なデータベースは最近一年分のみの提供。
  <ul>
    <li><a href="z39.50s://Tikal.dev.oclc.org/Medline">MEDLINE</a><br>
      医学の全領域にわたる文献 DB。収録範囲は 9,000 雑誌以上。(OCLC, FirstSearch)
    <li><a href="z39.50s://Tikal.dev.oclc.org/AidsCancerResearch">
      Aids and Cancer research</a><br>
      エイズ・癌研究に関する文献 DB。収録範囲は 2,440 雑誌以上。(OCLC, FirstSearch)
    <li><a href="z39.50s://Tikal.dev.oclc.org/MDX_Health">MDX Health</a><br>
      医学・薬学・生理学文献の引用・抄録 DB。(OCLC, FirstSearch)
  </ul>
</body>
</html>

```

図 3.3 Z39.50 データベースを表す URL を用いた HTML の例

上から Z39.50 データベースをハイパーリンクで結んでいる。そのため、WWW 上のどこからでも Z39.50 データベースを参照できる。また、標準化された形式 (URL) でデータベース接続情報が記述されているため、他の機関や個人が収集したデータベース接続情報を取り出し、自分の持つデータベースリストに追加することも容易である。

本研究では Z39.50 データベース接続情報を表す URL として RFC[49] で提案されている Z39.50 のための URL を採用した。この URL の書式は

z39.50s://ホスト名[:ポート番号]/データベース名

であり、例えば図書館情報大学デジタル図書館 [24] の OPAC データベースであれば、

z39.50s://lib.ulis.ac.jp:210/opac

と表される。また、URL を用いてデータベース接続情報を記述した HTML の例を図 3.3 に示す。

そして、Z39.50 用 URL が扱える WWW ブラウザとこのブラウザから呼び出される Z39.50 クライアントを提供することにより、ユーザはネットサーフィンしながら、Z39.50 データベースへのリンクをたどるだけで、データベースに接続し、検索をできるようになった。

このように、本研究におけるデータベース選択支援環境では、Z39.50 用 URL を自由に編集することも、相互に利用することもできるため、ユーザ間でお互い得意な分野のデータベースだけを収集・組織化するなど、お互いに補完しあうことができる。

3.3.2 WWW ブラウザ

WWW ブラウザは、

- HTML ビューア

- URL ハンドラ

により構成される（図 3.4 参照）。

HTML ビューアは Tcl/Tk で書かれたフリーで提供されている HTML ビューア [62] を利用した。

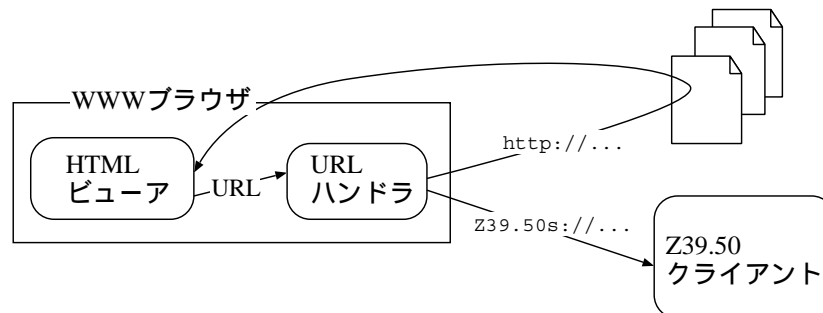


図 3.4 WWW ブラウザのシステム構成

本研究で開発した URL ハンドラは、`http://`や `z39.50s://`で始まる URL を解釈する。URL ハンドラは、HTTP 用 URL に対しては Tcl/Tk 付属の HTTP ライブラリを使い Web 上のファイルを取得し、HTML ビューアにそのファイルを渡す。Z39.50 用 URL に対しては、URL を引数として渡して Z39.50 クライアントを起動し、そのデータベースへ接続する。

本 WWW ブラウザの主な機能は以下の通りである。

- HTML2.0 相当へ対応：画像表示、ハイパーリンク表示
- HTTP 用 URL (`http://..`) に対応して WWW ドキュメントの取得、閲覧
- Z39.50 用 URL (`Z39.50s://..`) に対応して Z39.50 クライアントの起動
- HTML の `<a>` タグで示されたアンカーをクリックすることにより、リンク先を参照

このように、本研究では WWW ブラウザの基本機能と Z39.50 用 URL に対応する機能を実装した。なお、WWW ブラウザの大きさは 84KB である。

3.3.3 Z39.50 クライアント

本研究で開発した Z39.50 クライアントは、文献 [10] で発表した日本語化 Z39.50 クライアントに改良を加え、Z39.50 データベースを表す URL をコマンドライン引数で与えると、データベースに接続した画面を起動できるようにしたものである。この改良によって、WWW ブラウザからの Z39.50 クライアントの起動を実現した（実行例：3.3.4 節参照）。この Z39.50 クライアントは WWW ブラウザから呼び出されて起動するだけでなく、単体の Z39.50 クライアントとしても利用できる。本検索クライアントは著者や書名などのアクセスポイントを指定したり、前方一致や後方

一致、履歴を利用して検索できる。また、レコード数、履歴集合、詳細度合を指定して検索結果を表示できる。

3.3.4 実行例

我々が開発した WWW ブラウザは、既存の WWW ブラウザと同様にハイパーリンクをクリックしたり、URL を入力エントリー欄に直接入力することにより、WWW 上のコンテンツを閲覧できる。図 3.5 はウェブページ <http://cosmo.ulis.ac.jp/~yuka/wkappa/index.html> を表示したところである。

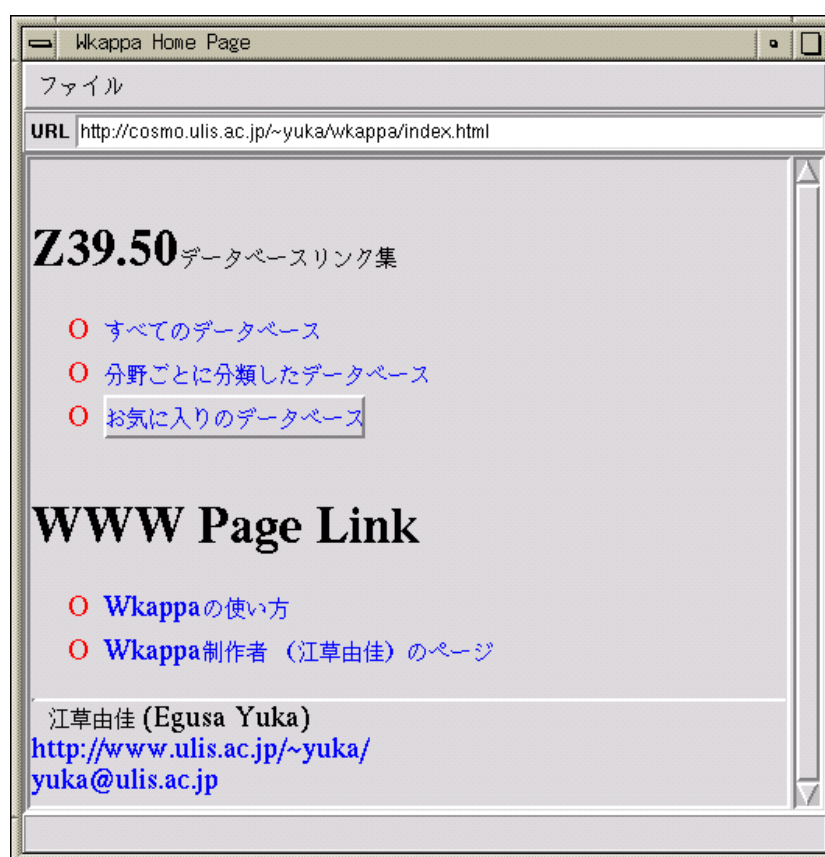


図 3.5 WWW ブラウザ

この WWW ブラウザは Z39.50 データベースを表す URL を解釈する。利用者が Z39.50 データベースへのハイパーリンクをクリックすると、Z39.50 クライアントが呼び出され、自動的にそのデータベースに接続するため、利用者はすぐに検索に移れる。図 3.6 はお気に入りのデータベースリンク集の中の「ULIS OPAC」(図書館情報大学のデジタル図書館 OPAC) のハイパーリンクを

クリックした結果、そのデータベースを検索できる Z39.50 クライアントのウインドウが、最前面に出ているところを示している。

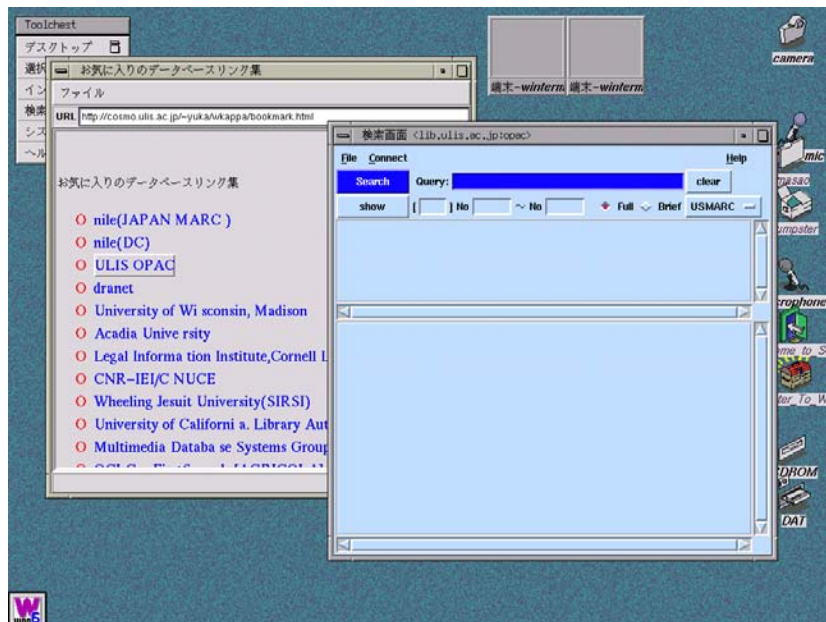


図 3.6 Z39.50 クライアントの起動画面

また、Z39.50 クライアントを起動したままの状態でも WWW ブラウザを利用できる。図 3.7 は複数のデータベースを検索しながら、本システムの使い方を説明しているウェブページを閲覧しているところである。このように、検索を行うだけでなく、関連するドキュメントの閲覧も同時に行える。

また、他の Z39.50 データベースを検索するために、次々と Z39.50 クライアントを呼び出すこともできる。図 3.8 は図書館情報大学の OPAC と我々が開発した Japan/MARC データベース [10] を検索しているところである。

図 3.9 は Japan/MARC データベース [10] のリンクをたどり、Z39.50 クライアントを使って検索しているところである。

検索履歴を順に見れば、

1. 「ガラス」を検索。（ヒットした件数は 667 件）
2. 「工芸」を検索。（ヒットした件数は 1373 件）
3. 1 番目と 2 番目の検索結果集合の AND 検索。（ヒットした件数は 19 件）

という検索要求が出されていることがわかる。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 をとることを示す。返戻結果表示

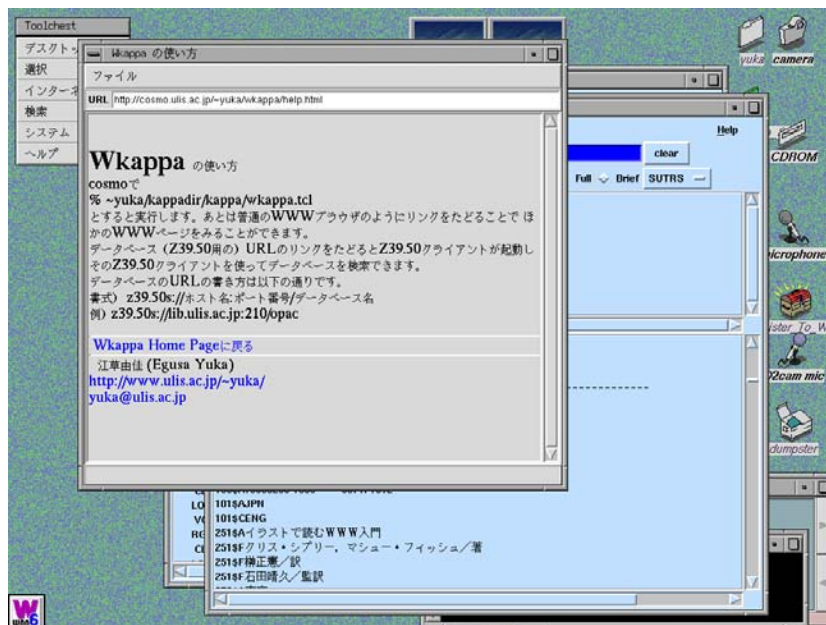


図 3.7 ドキュメント表示例

部分は 3 番目の検索結果集合を表示している。このように、Z39.50 クライアントは日本語の表示、入力、アクセスポイントを指定した検索、履歴集合を使った検索が行なえる。

3.3.5 開発環境

WWW-Z39.50 クライアントは Tcl/Tk と C 言語で開発した。動作確認は HP、SGI、Sun などの UNIX ワークステーション上の日本語化 Tcl/Tk8.0.5 で行った。WWW ブラウザからの Z39.50 クライアントの起動には Tcl/Tk の exec コマンドを用いた。また、Z39.50 プロトコルの解析は Index Data 社の YAZ (Z39.50 ライブラリ) [37] を利用した。

3.4 データベースリンク集の作成

データベース接続情報とは、データベースに接続するために、最低限必要な情報であり、具体的には、「ホスト名」「ポート番号」「データベース名」である。本研究ではこのデータベース接続情報を表すために RFC で提案された Z39.50 用 URL を用いた。この Z39.50 用 URL を利用したデータベースリンク集を公開しているサイトは見当たらなかったため、本研究では既に収集し、公開されている既存のデータベース接続情報を Z39.50 用 URL に変換しデータベースリンク集を作成した (図 3.10 参照)。

データベース接続情報を含めたデータベースに関する情報を収集しているサイトには、例えば、

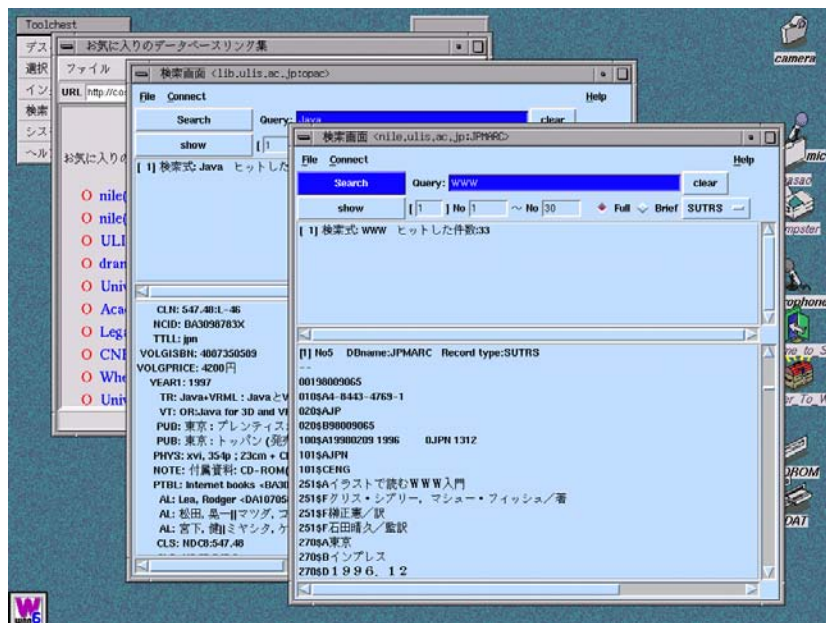


図 3.8 複数データベース検索実行例

英国の Z39.50 ターゲットに関する情報を提供するためのディレクトリサービス [63]、Index Data 社が作成しているデータベースに関する情報を収集しているサイト [20] などがあるが、本研究では後者から情報を取り寄せた。このサイトではデータベース接続情報だけでなく、そのほかのデータベースに関する情報についても収集している。例えばタイトル、著者名などアクセスポイントの種類、US-MARC やブレインテキストなど扱っているレコードフォーマットなどである。このサイトの対象は世界中で無料公開されているデータベースであり、収集データは 3 日に一度程度の頻度で更新されている。本研究では、このサイトから取り寄せた WWW ページから、データベース接続情報を抽出し、Z39.50 用 URL に自動変換するプログラムを作成した。

また、データベースを提供している組織自身が公開しているサイトなど、他の情報源から筆者自身が人手で収集し、Z39.50 用 URL で記述した。

こうして約 400 のデータベースの Z39.50 用 URL を作成した。

さらに、この Z39.50 用 URL を検索利用者が選択しやすいように組織化も行い、以下の 3 つのデータベースリンク集を作成した。

1. ホスト名ごとに並び替えた全データベースのリンク集：プログラムで自動的に並び替えて作成
2. 分野毎に分類したデータベースのリンク集：1 で作成したリンク集の中から、複数の分野の主なデータベースを抜き出し人手で分類して作成

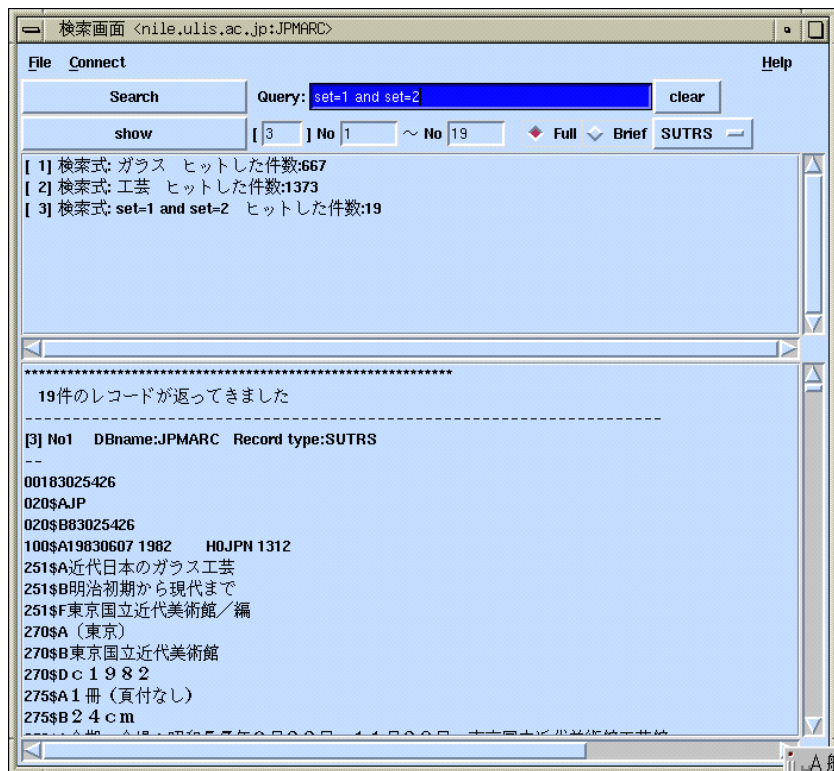


図 3.9 Z39.50 クライアントの検索画面

3. お気に入りデータベースのリンク集：1 で作成したリンク集の中から筆者がよく使うものを
人手で抜き出して作成

3.5 考察

これまで述べてきたように、本研究ではデータベース接続情報を共有でき、共有された情報を利用してすぐに検索できるデータベース選択支援環境を提案し、実装した。

この節では既存のデータベース選択支援とシステム構成の二つの観点から、関連システムと本研究との相違点を述べ、本研究のアプローチについて考察する。

3.5.1 directory-of-servers との比較

本研究とは異なるアプローチでデータベース選択支援機能を実現しているものに WAIS(Wide Area Information Servers) の directory-of-servers がある。WAIS では、データベースに関する情報を扱う特別なデータベース directory-of-servers を設け、利用者がこのデータベースを検索すること

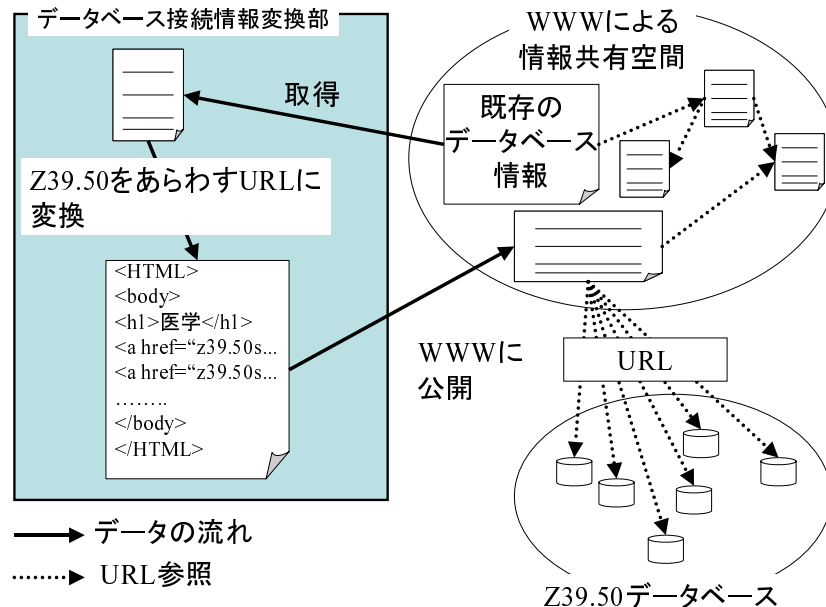


図 3.10 データベースリンク集の作成

により、目的のデータベースを見つけ、データベースに接続するための情報を知ることができる[64]。しかし、利用者がデータベースを見つける手段は、directory-of-servers を検索するしかない。

WAIS のアプローチと本研究のアプローチではデータベース接続情報を利用者側から発信できるかどうか異なる。WAIS では directory-of-servers という特別なデータベースを作成しなければならないが、このデータベースを作成し更新していくのは簡単ではない。しかし、本研究ではデータベース接続情報を記述したデータベースリンク集を作成して WWW 上に置くだけで良く、利用者の側からも容易にデータベース接続情報を発信できる。

なお、Z39.50 においても、2000 年 1 月に行われた Z39.50 Implementors Group (ZIG) Meeting において将来必要な機能として、同じような機能をもつ Distributed Directory of Z39.50 Servers[65] が議論され始めたが実装はされていない。

3.5.2 Z39.50-WWW ゲートウェイとの比較

directory-of-servers の他に、データベースを分野ごとに分類して、WWW-Z39.50 ゲートウェイシステムで提供しているところもある。WWW-Z39.50 ゲートウェイは WWW 上からデータベースを検索できる点が本環境と似ているが、WWW-Z39.50 ゲートウェイシステムは、本環境とはシステム構成が異なる。

WWW-Z39.50 ゲートウェイシステムでは、ユーザからの検索要求は HTTP プロトコルを介し

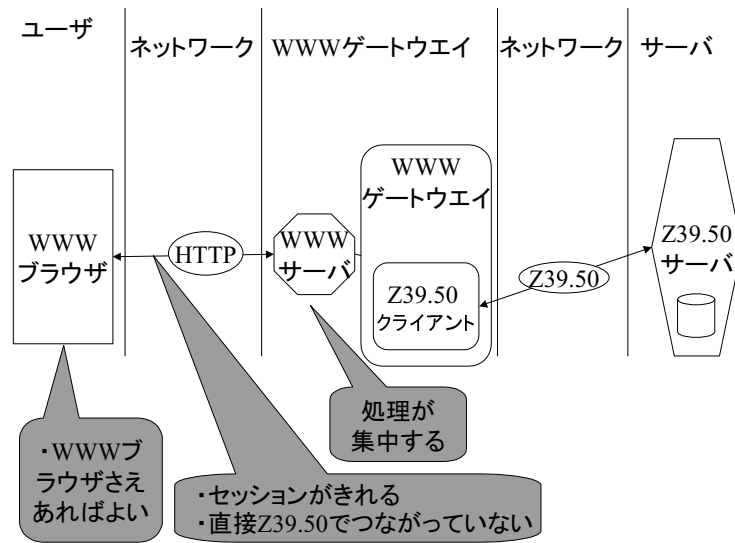


図 3.11 WWW-Z39.50 ゲートウェイシステムとプロトコルとの関係

て、いったん WWW サーバに送られ、WWW ゲートウェイが Z39.50 プロトコルに変換して Z39.50 サーバに接続する。また、Z39.50 サーバからの返信は、WWW ゲートウェイが Z39.50 を解析し、HTTP を介してユーザに送られる (図 3.11 参照)。

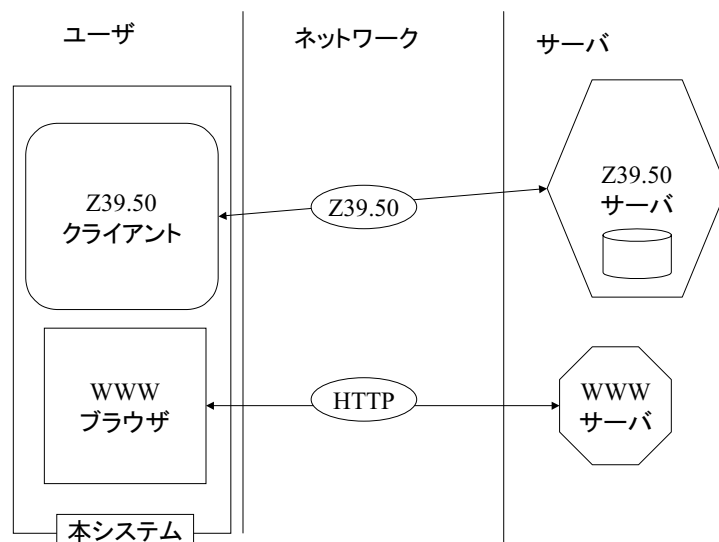


図 3.12 本システムのプロトコルとの関係

これに対して、本環境では、情報検索には Z39.50 を使い、情報共有には WWW を使うというように、Z39.50 と WWW をそれぞれ適した役割でプロトコルを使い分けている（図 3.12 参照）。

WWW-Z39.50 ゲートウェイシステムでは、ユーザ側は WWW ブラウザのみで検索できるため、多くの機関で採用されているが、HTTP プロトコルから Z39.50 プロトコルへの変換を行っているため、以下の問題がある。

- 処理が複雑になる
- 全てのユーザの検索処理は WWW サーバに集中してしまう

そのため、WWW-Z39.50 ゲートウェイシステムでは Z39.50 を利用する利点の一つである利用者が使い慣れたユーザインタフェースで検索することができず、分散型検索システムの特徴の一つである処理の分散化が損なわれる。

また、WWW-Z39.50 ゲートウェイシステムでは、新たなデータベースを追加したり、組織化したりできるのは、WWW-Z39.50 ゲートウェイシステムの管理者だけであり、利用者個々人が自由にカスタマイズできず、利用者同士でのデータベースに関する情報の共有もできない。

3.5.3 その他の関連研究

データベース選択支援という観点ではないが、本研究に関連した研究について述べる。

まず、プロトコルを検索と情報共有において使い分けるという点で本研究と類似している研究について述べる。

Hammer ら [66] は Z39.50 と WWW を組み合わせたシステムが良い結果をもたらす可能性を指摘したが、具体的な方策は述べていない。

また、WWW 上から Z39.50 データベース資源を利用できるシステムとして、Z.X[67] が最近発表された。Z.X では WWW ブラウザ用に Z39.50 と XML の間のプロトコル変換を行うハンドラを開発（2000 年 7 月現在ベータ版）し、Web のユーザインタフェースで Z39.50 サーバを検索できるようにしつつある。Z.X はデータベースから返されるレコードを XML で共有することに主眼をおいたシステムであり、本研究のようにデータベース選択のための情報を WWW で共有し、ユーザの選択支援を行う環境については触れていない。

2.4.2 節で述べたように、様々なプロファイルが存在する場合、データベースを選択する際に、どのプロファイルを採用しているかが分かることは有用である。本研究においては、プロファイルの情報は、Z39.50 データベースリンク集にプロファイルの情報を記述することで、共有が可能である。

また、Z39.50 データベース全体の選択支援をどのように行うかについての言及がなされたプロファイルは見当たらない。

3.6 まとめ

本研究では以下のアプローチで、データベース選択支援のための環境を提案するだけでなく、その環境を実践できるシステムも作成した。

1. Z39.50 と WWW をそれぞれ適した役割に使い分けたシンプルな構成
2. データベースリンク集を作成すればよいというユーザからの情報発信の容易さ
3. グラフィカルな WWW-Z39.50 クライアントの開発

本研究を元にした選択支援環境が整備されることにより、分散型検索システムについての情報共有ができ、情報検索の向上につながると筆者は考える。

本研究では、検索利用者のデータベース選択支援を目的として、データベース接続情報を共有でき、リンクをたどるだけで直接検索できる環境を提案し実装した。本研究によって、単独のクライアントと比較して、データベースを選択するための手段が増えたのみならず、Z39.50 データベース接続情報の利用者間での交換や共有が可能になった。

従って、3.2 章に述べたように、利用者は、各自の好みに合ったデータベースリンク集を任意に作成し利用できるようになった。さらに、利用者は、ある分野に精通した人、例えば図書館員やデータベース提供者などが作成したデータベースリンク集も利用できるようになった。その結果、利用者の検索要求に適合したデータベースを選択しやすくなった。

第 4 章

分散環境におけるデータベースに関する情報の共有システム

4.1 はじめに

現在、インターネットと World Wide Web (WWW) の普及にともない多種多様なデータベースサービスが提供されるようになり、場所、時間を問わず利用可能なデータベースは増え続けている。その上、これらの中には無料で提供されているものも多く、以前より、手軽に多くのデータベースを利用できるようになった。一方で、利用者による検索行動も多様化している。これは、利用者自身の多様化によるところが大きい。つまり、WWW 以前には専門家が担ってきた検索はエンドユーザの手に移り、児童から主婦・老人と、老若男女や職業を問わず、様々な種類のユーザが検索システムを利用するようになった。

そのため、利用可能なデータベースを把握すること、検索要求に合ったデータベースを適切に選択すること、データベースを使いこなすことは、困難となっている。

これらの問題に対応するために、データベースに関する情報をまとめたデータベースディレクトリやノウハウ本などが出ているが、これらに載っているのは、編纂者の把握している情報のみであり、日々増加している新たなデータベースの発見・把握という問題に対処できない。さらに、データベースを使いこなす際に必要な利用の際のノウハウについても編纂者の固有の観点からまとめられているため、多様なユーザのデータベース利用に対応できない場合も多い。

そこで、我々はこれらの問題を解決するために、データベースに関する情報を利用者同士で共同入力・共同利用可能にすることを提案する。我々はこれまで、Z39.50 書誌検索システムを対象としたデータベース選択支援環境 [11] の構築を行ってきたが、本研究では、対象を Z39.50 データベースに限らず、他のデータベース全般に広げた。さらに、本研究では、データベースの基本情報を設計し、その基本情報を入力・閲覧・再利用でき、また、基本情報だけでなく、ユーザ毎のノウハウなども入力・閲覧できるシステムを作成した。

本章では、データベースの把握・選択・利用の問題を扱った関連研究、データベースに関する情報とその共有についての本研究のアプローチ、構築したシステム、本研究と類似したアプローチの研究との類似点・相異点、本システムの運用方法や適用例について述べる。

4.2 関連研究

データベースの把握・選択・利用の問題に対する本研究のアプローチは、データベースに関する情報を利用者同士で共同入力・共同利用可能とすることである。これまでデータベースの把握・選択・利用の問題に対して同様のアプローチで解決を図ったものは見あたらなかった。本節では、この問題を扱った関連研究について述べる。

データベースに関する基本的な情報をまとめて検索できるようにしたものにデータベースディレクトリがある。日本では、データベース台帳 [68][69]、国立情報学研究所のデータベースディレクトリ [70]、世界 CD-ROM 総覧 [71] があり、海外では Gale Directory of Databases[72] などがある。また、WWW 上のデータベースを対象とした Dnavi[73] もある。これらは、アンケートや報告書や編纂する組織自身の調査を元にデータベースの情報を収集し検索できるようにしたものである。これらのデータベースディレクトリを検索することで、データベースに関する基本的な情報を得ることができる。

しかし、これらは、収録範囲が限定されている (商用のもののみ、CD-ROM で提供されているもののみなど) ことや、更新が遅い (年に 1、2 度程度) という問題がある。また、掲載されているのが基本情報のみであり、具体的な検索方法などのノウハウは得られないという問題もある。

情報検索のための入門書や、分野ごとのノウハウ本 [53][54][55] などのように、代表的なデータベースにしばってまとめ、具体的な利用方法を説明しているものがある。これらの本は様々な目的のために作成された本であり、その本の筆者の観点で書かれており、網羅性に欠けるという問題がある。また、日々変化するデータベースの情報の更新ができない。

また、WWW が普及する以前には、エキスパートシステムを使ってデータベース選択を支援する研究もいくつか行なわれている [74][75]。これらは検索の専門家のデータベース選択における戦略をモデル化し、エキスパートシステムとして構築したものである。データベース選択を行なう際の情報については冊子媒体のデータベースディレクトリや検索の専門家による知識を用いている。本研究との違いは、エキスパートシステム運営者が把握しているデータベースについてのみが対象となる点であり、新たなデータベースの発見や把握については、エキスパートシステム運営者に依存する。

データベースに関する情報を提供する機能を備えた既存のシステムに、WAIS[76] の directory-of-servers がある。WAIS では分散して存在するデータベースについて、その所在とそのデータベースに関する説明やキーワードなどをまとめたものを用意し、それらを directory-of-servers に載せる。所在や説明などのデータベースの基本情報を、共有する観点から本研究と類似している

が、本研究とは異なり、データベースの提供者のみが、基本情報を directory-of-servers に登録でき、利用者同士での共同入力の観点はない。対象についても異なり、本研究ではデータベース全般を扱っているが、directory-of-servers では WAIS プロトコルに基づくデータベースのみである。また、ユーザのノウハウについても取り扱えない。

Meng ら [77] によれば、メタサーチエンジンにおけるデータベース選択の手法には、大きくわけて、人手により特徴を記述するもの、データベースに含まれる語についての統計的な情報を利用するもの、検索式に応じて学習を行なうものの 3 つがあり、それぞれ様々な手法が提案されている [78][79]。メタサーチエンジンにおけるデータベース選択とは、メタサーチエンジンに登録されているデータベース群という特定の集合の中からデータベースを選択することである。これとは異なり、本研究は分散環境で日々増えている未知のデータベースを含む全データベースを対象としている。このことはデータベースの発見・把握という観点で大きく異なる。

また、岩澤ら [80] によるレファレンスのノウハウを共有することを目的とした研究もある。これは、レファレンスにおける質問応答の内容を組織化し共有するものである。この研究では、質問応答において使われたレファレンスツールの情報を入力し共有する点が本研究と類似しているが、岩澤らの研究の主眼は、レファレンスにおけるノウハウの共有であり、基本情報の再利用の観点はない。

また、コレクションのメタデータを設計し、共有するという研究 [81][82] もある。ここでいうコレクションとは、図書館や博物館のコレクションや、Yahoo などのディレクトリ型サーチエンジン、Alta Vista などの全文サーチエンジン、デジタルアーカイブなど様々なものを含むものであり、この中にはデータベースも含まれる。この研究では、コレクションの基本情報を共同入力・共同利用する点が本研究と類似しているが、コレクションメタデータの対象は静的なコレクションを含むもっと広いものであり、また、利用者のノウハウを共同入力・共同利用しようという観点はない。

4.3 データベースに関する情報の共有

データベースに関する情報の共有とは、情報を作成したユーザ自身のみがその情報を利用できるのではなく、作成者以外のユーザもその情報を利用できることである。例えば、各データベースの概要がわかる情報や、複数のデータベースをある主題でまとめた情報などをユーザが自身のために作成したとしても、他のユーザがそれらの情報を閲覧してデータベース利用に活用したり、さらにコメントなどの情報を追加したりできることを指す。

本研究で提案するデータベースに関する情報共有システムでは、利用者同士がデータベースに関する様々な情報を共有できる。本研究ではデータベースに関する情報を基本情報と付加情報に分け整理した。これらについては 4.3.1 節と 4.3.2 節で詳しく述べる。

図 4.1 にデータベースに関する情報の共有システムによる情報共有のイメージ図を示す。利用者

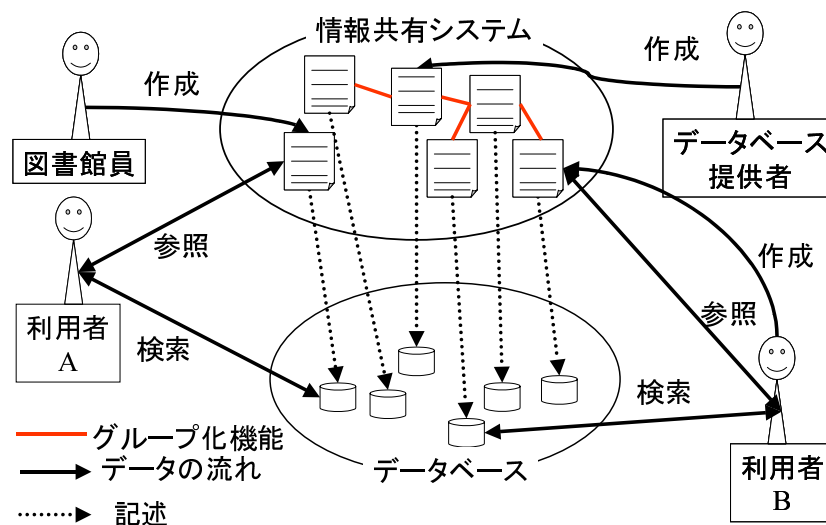


図 4.1 データベースに関する情報の共有イメージ

は、共有した情報をもとに検索要求にあったデータベースをみつけたり、データベースの使いかたを知ったり、データベースを自分なりにまとめたりできる。

共有システム上においてどのような利用者がどのような動機で情報を入力するかは様々である。以下にその代表例を挙げる。

- 図書館員が、自館にてサービスを提供しているデータベースのリストを自館の利用者むけに作成する
- データベースの提供者が、自らの提供するデータベースを利用者に広報し、利用促進を図る目的で、データベースの概要がわかる情報を入力する
- 一般利用者が、自身の利用に便利なデータベースについて、自身が以後利用する際の覚え書きとして、データベースのリストを作成する
- ある主題に詳しい人が、その主題を扱っているデータベースについて、それらのデータベースの特徴がわかる情報を入力したり、主要なデータベースをピックアップしてリストを作成したりする。

利用者は、既にシステム内に存在する基本情報を参照しながら、望みのデータベースに接続して検索を行ったり、また自分の要求に合ったデータベースを探したりできる。例えば、図 4.1 において、利用者 A は図書館員の作成した情報を参照して、データベースに接続して検索している。利用者 B は自身が作成した情報を参照して、ある観点でグループ化されたデータベースについての情報を辿って、データベースに接続して検索している。

このように、情報を相互に利用できるため、情報を誰かが入力すれば、他の利用者が新しいデー

データベースの存在を知ることができる。また、データベースではデータが同じで、検索システムが異なる場合も多い。例えば、The Institution of Electrical Engineers(IEE:英国電気学会) が作成している「INSPEC」というデータの場合、「OVID 検索システム」,「アット・ニフティ」,「DataStar」,「DIALOG」,「STN International」,「JICST オンライン情報システム (JOIS)」などの様々な検索システムがある。そのような場合でも、既に誰かが入力して、システムに存在している情報を取り込んで、流用し、必要な部分だけと修正して新たに入力することができる。例えば、上記の例の「INSPEC」というデータを使った新たなシステムを追加する場合であれば、入力者は、「OVID 検索システム」の「INSPEC」データベースを取りこんで、データ「INSPEC」の特徴や収録している分野やデータの作成者などのデータに関する記述はそのまま流用し、検索システムの名前「OVID 検索システム」や利用条件、更新頻度などの変更部分だけを修正して入力することができる。

さらに、入力された情報を媒体として、ユーザ同士がコメントを入力しあったり、複数のデータベースを関連づけることによって、各ユーザが持っているノウハウをも共有することができる。

4.3.1 データベースの基本情報

データベースに関する情報には、様々なものがあるが、その中でも大抵のデータベースに共通しており、データベースの概要を把握できる基本的な情報がある。例えばデータベースの名前、システムの名前、データベース作成者の名前、データベースの収録範囲、収録数、収録言語、扱っている分野、利用条件などである。本研究ではこれを基本情報と呼ぶ。

基本情報は客観的な情報であり、入力者の観点によって変わることが少なく、データとして再利用しやすい情報である。そこで本研究では、基本情報はデータ単体として再利用しやすいように、記述する項目を選別し XML[83] として記述可能にした。

設計した基本情報の項目説明を表 4.1 に示す。設計した基本情報の項目の選定にあたっては、データベースの概要が把握でき、データベース選択の際に利用者が利用コストや検索要求に適しているかどうか判断する材料となる項目かどうかを考慮した。派生元 ID を作成した理由は、基本情報を取り込み入力した際に、元となる基本情報を基本情報に入れることで、引用元が明確になるためである。この項目のデータを使うことで、システムで引用元に辿ることができる。またよく引用されるデータベースがなにかを統計的に調べることも可能となる。登録者 ID を基本情報に入力する理由は、基本情報は入力者自身のみが更新可能とするためである。実際に作成した基本情報の例を図 4.2 に示す。このデータ例は、データベース台帳総覧 [68] をもとに作成したものである。

4.3.2 データベースの付加情報

データベースに関する情報には、基本情報以外にもある。実際に利用した際に注意した点や、検索のためのちょっとしたコツや、複数のデータベースをさまざまな観点でグループ化した情報など

表 4.1 基本情報の項目説明

項目名	タグ名	説明
ID	<id>	基本情報の ID を表す。
派生元 ID	<source_id>	基本情報を取り込み入力したときに、取り込んだ元の基本情報の ID を表す。
作成日	<created_date>	基本情報を作成した日時を表す。システムが自動的に挿入。
更新日	<update_date>	基本情報を更新した日時を表す。システムが自動的に挿入
登録者 ID	<userid>	基本情報を登録した人の ID を表す。入力・閲覧システムのアカウントでもある。システムが自動的に挿入
データベース名	<dbname>	データベース名を表す。必須入力項目である。
システム名	<system>	システム名を表す。
利用条件	<condition>	利用条件を表す。利用時間、利用料金などを含む。
データ提供形態	<format>	データの提供形態を表す。オンライン、MT、CD-ROM など。複数可。
コントリビュータ名	<contributor>	データベースの寄与者をあらわす。データの作成者（プロデューサ）やシステムの構築者、システムのサービス提供者（ディストリビュータ）などがある。複数可。
データベースの特徴	<description>	データベースの特徴の説明を表す。データの特徴やシステムの特徴など。
主題	<subject>	データベースの主題をあらわす。分野やキーワードなど。複数可。
データのタイプ	<type>	データのタイプを表す。文書、写真、図表などがある。複数可。
データの記述言語	<lang>	データの記述言語を表す。複数可。
収録期間	<period>	データの収録期間を表す。
収録件数	<total>	データの収録件数を表す。
更新頻度	<interval>	データの更新頻度を表す。
更新件数	<interval_num>	データの更新件数を表す。
収録情報の地域	<region>	収録情報の地域を表す。
カテゴリ	<category>	システムが自動的につけるカテゴリを表す。
アクセス先	<access>	アクセス先を表す。

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE database_metadata SYSTEM "db.dtd">
<database_metadata>
  <id>9354</id>
  <source_id/>
  <created_date>2002-06-20</created_date>
  <update_date>2002-06-20</update_date>
  <userid>0177</userid>
  <dbname>テレアド</dbname>
  <system>アドレスナビプラス</system>
  <condition>サービス料金： 地域内容による
</condition>
  <format>パッチ (CD)</format>
  <contributor role="producer">(株) ダイケイ
</contributor>
  <contributor role="distributor">(株) ダイケイ
</contributor>
  <description>電話帳データベース「テレデータ」と
(株) ゼンリンの住宅地図表札データとが結合された、
統合データベース。電話帳データベースではカバー
できない情報もテレアドにより統一されたデータベース
として作成されている。6,000 万件のテレアドがパソ
コンにより簡単に検索できる。</description>
  <subject>地名</subject>
  <subject>地図</subject>
  <subject>住所</subject>
  <subject>電話番号</subject>
  <subject>住所</subject>
  <subject>氏名</subject>
  <type>文書 (全文)</type>
  <lang>日本語</lang>
  <period>1999 年 ~</period>
  <total>6,000 万件</total>
  <interval>年 1 回</interval>
  <interval_num/>
  <region>日本</region>
  <category/>
  <access/>
</database_metadata>

```

図 4.2 データ例

がある。本研究ではこのような情報を付加情報と呼ぶ。

付加情報は、データベースの種類や対象領域によって大きく異なり、自由記入文として入力されることが想定されるため、再利用するための一定のフォーマットとして一律に記述することが困難であったり、再利用するときに大幅な変更が必要になる情報である。

付加情報には有用なものも多く、共有することでデータベースを利用するのに役立つ。しかし、付加情報だけではデータベースは利用できず、基本情報を得る必要がある。このため、付加情報と基本情報とを組み合わせると参照できると効果的である。そこで、本研究では付加情報と基本情報を組み合わせると入力・参照できるようにした。

具体的には、本システムのグループ化機能を使うことにより、主題ごとにデータベースをまとめたり、同じようなデータベースをまとめることができる。また、データベースを様々な観点でまと

められ、そのまとめられたデータベースを基本情報とともに見比べることができる。そして、各データベースの基本情報を見ながら、どの利用者でもそのデータベースのコメントが入力できるため、ちょっとしたコツや、疑問や答えなども利用者同士のやりとりで得ることができる。

4.3.3 サンプルデータの作成

本システムに入力したサンプルデータは、著者自身が調査して一から作成したものと、「平成 12 年度データベース台帳総覧電子化版」(以下データベース台帳総覧と呼ぶ)を元に作成したものがあ
る。ここでは、データベース台帳総覧から作成したサンプルデータについて述べる。

データベース台帳総覧は、法律に基づいてデータベースサービス企業から経済産業省に提出された申告書をまとめて作成されたものである。そのため、主に商用目的のデータベースに関する情報が収録されている。

データベース台帳総覧には以下の 4 つのファイルがある。

- 企業ファイル 208 件
申告企業の企業名、住所などが収録されている。
- データベースファイル 3614 件
データベース名、プロデューサ名、分野、キーワードなどデータやデータベースに関する情報が収録されている。
- システムファイル 3658 件
システム名、ディストリビュータ名、使用条件などシステムに関する情報が収録されている。
- ダブリファイル 562 件
データベース台帳総覧のデータは、異なる企業が同じデータベースを申告したり、データが同じでシステムが異なる複数のデータベースが申告されたとしても、それら申告された記述を修正したり、どちらか一方を採用したり、記述内容をまとめるなどは行っていない。その代りに、経済産業省が同じデータベースと思われると判断したデータベースの集合を列挙したダブルファイルを用意している。
同じと思われるデータベースの ID が一行に「,(カンマ)」区切りで並べられている。例を以下に示す。

1865,4227

1089,1866,4228

この例の場合、データベース ID1865 と 4227 の 2 つが同じと思われるデータベースの集合であることを表し、データベース ID1089 と 1866 と 4228 の 3 つが同じデータベースと思

われる集合であることを表している。

本システムのサンプルデータとして利用するために、著者は以下の方針でデータベース台帳のデータを変換した。

- 企業データを登録者データとして扱う

データベース台帳は、各企業から申告されたデータから作成されているため、各企業は本システムでいうところの基本情報登録者といえる。そこで、本研究では、各企業を基本情報登録者と見立て、企業データを登録者データとして扱った。こうすることによって、本システムに、複数の登録者が登録できることを示した。

- ダブリファイルをグループ情報として扱う

ダブリファイルは、同じデータベースと思われるものが収録されているため、同じデータベースであるという観点で複数のデータベースをグループ化しているといえる。そこで、本研究では、ダブリファイルでグループ化されている複数のデータベースを、本システムのグループ化機能のグループと見立てる。こうすることによって、本システムにおいて、複数のデータベースを関連づけるグループ化が可能であることを示した。

著者が設計した基本情報とデータベース台帳との対応を表 4.2 に示す。

4.4 データベースに関する情報共有システム

本研究では、前節で述べたデータベースに関する情報を共有するためのシステムを開発した。本節ではこの共有システムについて述べる。

システムの機能は以下の通りである。

- ブラウジング
- 検索（フィールドごと、全文）
- 基本情報登録・更新（ユーザによる）
- コメント機能
- グループ化機能
- ユーザ認証

各機能の詳細は 4.4.2 節以降にて述べる。

4.4.1 システムの構成

図 4.3 に本システムの構成を示す。登録された基本情報は XML ファイルとして蓄積され、検索や一覧表示のアクセス用に、項目ごとのインデックスがそこから構築される。付加情報については

表 4.2 基本情報とデータベース台帳の項目との対応表

項目名	データベース台帳
ID	データベース NO.
派生元 ID	なし
作成日	なし
更新日	なし
登録者 ID	企業 NO.
データベース名	データベース名
システム名	システム名
利用条件	サービス料金、サービス時間帯、使用条件
データ提供形態	データ提供形態
コントリビュータ名	ディストリビュータ、プロデューサ
データベースの特徴	データベースの特徴、システムのコメント
主題	分野、キーワード
データのタイプ	データのタイプ
データの記述言語	データの記述言語
収録期間	収録期間
収録件数	収録件数
更新周期	更新周期
更新件数	更新件数
収録情報の地域	収録情報の地域
カテゴリ	なし
アクセス先	なし

ブレインテキスト形式でファイルに登録される。検索やブラウジングの際に CGI プログラムはこれらのデータファイルにアクセスし実際の検索や表示を行なう。特に、基本情報の XML データの表示の際には XML ファイルに対して WWW サーバ内で、XSLT スタイルファイルを適用しその結果をブラウザに渡す。

本システムは、CGI の枠組みを利用して Web 上でユーザの要求に応じて基本情報の閲覧・登録やグループ情報の登録・コメントの登録ができるシステムとして構築した。開発に利用した環境は以下の通りである。WWW サーバには Apache [84] を使い、ユーザ認証には WWW サーバの Basic 認証機構を利用した。CGI プログラムは Perl5 [85] で作成した。項目ごとの検索などを高速に行うために Berkeley DB ライブラリ [86] を利用したインデックスを作成した。また、XML から

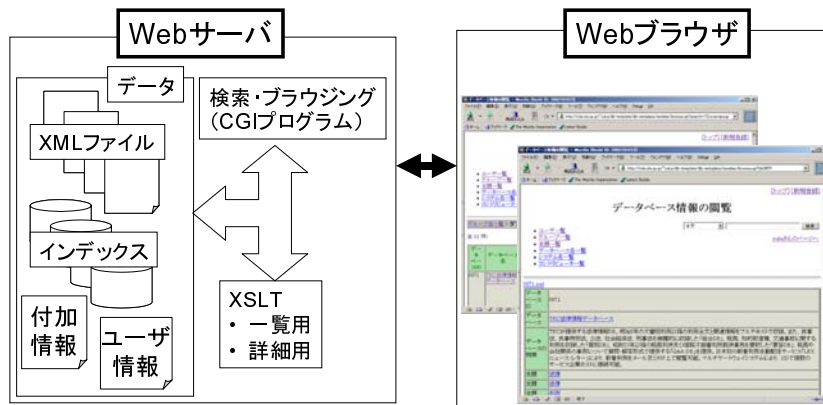


図 4.3 システムの構成図

各種表示形式への変換を行う XSLT 処理系には Libxslt [87] を利用した。

4.4.2 ブラウジング機能

ブラウジング機能は、データベース情報についての簡易表示、詳細表示、一覧表示、ユーザページ表示を中心として、データベースの基本情報や付加情報、ユーザ情報などを表示して、ユーザに提示する。

簡易表示

簡易表示では、20 件ずつ各データベースのデータベース ID、データベース名、特徴、主題を表示する。後述する検索機能などで複数のデータベースを表示する場合などにも、簡易表示機能を用いる。

図 4.4 は、システムの初期画面で使われている簡易表示である。画面中央の表は、一行が一つのデータベースの基本情報を表しており、登録された順に最新のもののほど上に表示される。表の 2 行目を説明すると、データベース ID が「9388」、データベース名が「JAPAN/MARC 検索システム」、概要が「Z39.50-JAPAN/MARC 検索システムに使ったサーチエンジンを... (以下省略)」、主題が「全般 (書誌情報); 出版:... (以下省略)」の基本情報を表していることがわかる。これは筆者が作成した基本情報のサンプルである。

データベース名のリンク、例えば「JAPAN/MARC 検索システム」を辿ることにより基本情報の詳細表示を見ることができる。

図 4.4 の画面下部の「ページ：」の右の数字のリンクを辿ることによって次の 20 件を見ていくことができる。

データベース情報の登録/更新 - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 進む 検索 お気に入り メディア

アドレス http://nileulis.ac.jp/~yuka/db-template/db-metadata/browse.cgi

データベース ID	データベース名	特徴	主題
9389	ハイパー薬事典	名前や記号などで薬を検索できる。検索結果で得られる情報は非常に詳しく、特にステロイドの場合、注意事項にきちんとステロイドのレベルが記述されているのはすば...	処方薬, 医学, 病院, 薬, 薬学;
9388	JAPAN/MARC検索システム	239.50-JAPAN/MARC検索システムに使用したサーチエンジンを実験的にWWW上で公開しています。現在、1,019,696 のレコード(1983年か...	全般(書誌情報); 出版; 図書・雑誌所蔵; 図書・雑誌目録;
9387	LISA Plus	図書館情報学の分野と出版業および図書館販売などのような関連諸領域の情報を収録。また、進行中もしくは完成した研究プロジェクトを国際的な範囲で収録した印刷体のCu...	図書館・情報学; 図書館情報学; 社会学;
9386	J-BISC	国立国会図書館の全国書誌情報。JAPAN-MARCのCD-ROM版。約130万件以上の書誌データを収録。図書カード形式の印刷や、ダウンロード機能など、簡単な操...	全般(書誌情報); 図書・雑誌所蔵;
9385	雑誌記事索引 遡及版 1975-1984	国内で刊行された学術雑誌約9,000誌の主要記事の書誌データを検索できるCD-ROMである。人文・社会及び医学、薬学を含む科学技術と幅広い分野から採録し、論題...	全般(書誌情報); 出版; 図書・雑誌所蔵; 図書・雑誌目録;
9384	雑誌記事索引 遡及版 1985-1994	国内で刊行された学術雑誌約9,000誌の主要記事の書誌データを検索できるCD-ROMである。人文・社会及び医学、薬学を含む科学技術と幅広い分野から採録し、論題...	全般(書誌情報); 出版; 図書・雑誌所蔵; 図書・雑誌目録;
9383	雑誌記事索引 カレント版	国内で刊行された学術雑誌約9,000誌の主要記事の書誌データを検索できるCD-ROMである。人文・社会及び医学、薬学を含む科学技術と幅広い分野から採録し、論題...	全般(書誌情報); 出版; 図書・雑誌所蔵; 図書・雑誌目録;
9382	ULIS OPAC	図書館情報大学で所蔵している蔵書の検索ができる。筑波大と図書館情報大学が統合したため、図書館情報大学の付属図書館だけを限定して検索することによって、検索でき...	図書館; 図書館情報学; 情報; 情報科学;
9377	英辞郎	日英・英日の辞書。非常に豊富な用例がよい。	全般; 英訳; 英語;
9376	BUNSOKU2003(管理・システム技術編)	管理分野では、経営工学のシステム工学やオペレーションズリサーチ、生産工学の品質管理や安全管理、情報管理の情報検索などを収録しています。システム分野では、計算...	
9374	Dnavi	WWW上にあるデータベースを検索できるデータベースのデータベース。	
9361	Google	サーチエンジンです。公開されているWebドキュメントを検索できます。PDFなども検索できる。	Webドキュメント; 全般;
9360	つくば市立中央図書館蔵書検索サービス	つくば市立中央図書館の蔵書が検索できます。	全般; 娯楽; 書誌; 生活;
9359	NACSIS Webcat	総合目録データベースWWW検索サービス。かなりの大学の蔵書が一度に検索できる。検索結果に所蔵一覧もあるので、どの大学に蔵書があるか調べるときに利用すると良...	全般; 書誌;
9358	Vector	豊富なフリーソフトを検索できる。gooで検索を調べると、検索結果の簡易表示に、ソフトがフリーかどうか一目瞭然とわかるので便利。	シェアウェア; フリーソフト;
9357	The Asia-Pacific Collection	「米国議会委員会が発行する公聴会議事録、レポート、配布資料からなるSerial Setのうち、1970年以降1997年までの期間で、米国によるAPEC加盟諸国...	APEC; アジア・太平洋政策;
9356	筑波大学蔵書検索	OPAC。筑波大学で所蔵している蔵書の検索ができる。	OPAC;
9355	NDL-OPAC	国立国会図書館の蔵書検索と蔵書申し込みができる。	全般; 書誌;
9354	テレアド	電話番号データベース「テレデータ」と(株)ゼンリンの住宅地図表札データとが結合された、統合データベース。電話番号データベースではカバーできない情報もテレアドにより...	住所; 地名; 地図; 氏名; 電話番号;
9353	Drug R&D Backgrounders	医薬品情報の新しい変化に対応するためインターネットベースの出版物として、独自分類の全36セクションにわたる疾病分野から構成。医化学文献、特許文献、学会情報の他...	化学; 化学構造・用語; 医学; 医薬; 特許; 生物学; 研究動向; 薬学;

ページ: [1] [2] [3] [4] [5] [6] [7] [8] [9] ... [182]

インターネット

図 4.4 最新基本情報簡易表示 (初期画面)

詳細表示

詳細表示では、データベースの基本情報の項目のすべてを表示すると共に、そのデータベースに付与されたコメントや、グループ情報も全て表示する。なお、基本情報のうち、キーワードと分野などの項目表示は、それぞれ同じキーワード、分野を持つデータベース検索へのリンクになる。

図 4.5 は JAPAN/MARC 検索システムの詳細表示である。左上部の「9382.xml」は、この基本情報の XML 表現へのリンクであり、ユーザはこの基本情報を再利用可能な XML データとして取り



図 4.5 基本情報詳細表示

寄せることができる。表は基本情報のうち入力されている全項目を表示している。左の列は項目名を右の列は実際のデータを表している。データベース、主題、システム、コントリビュータ各項目のデータはリンクになっており、そのリンクを辿ると、その項目に同一の内容を持つ基本情報の簡易表示を見ることができる。例えば主題の項目の「出版」のリンクを辿ると、主題が「出版」のデータベースが簡易表示される。

下部にはグループ Japan/MARC(M) の表示や、コメントの表示がある。これらは後の節のコメント機能、グループ化機能の所で説明する。

一覧表示

一覧表示では、基本情報の項目やログインユーザが同じ値を持つものをまとめてブラウジングできるように、「主題」「データベース」などの項目ごとに、その項目の値を出現頻度順に並べて表示する。

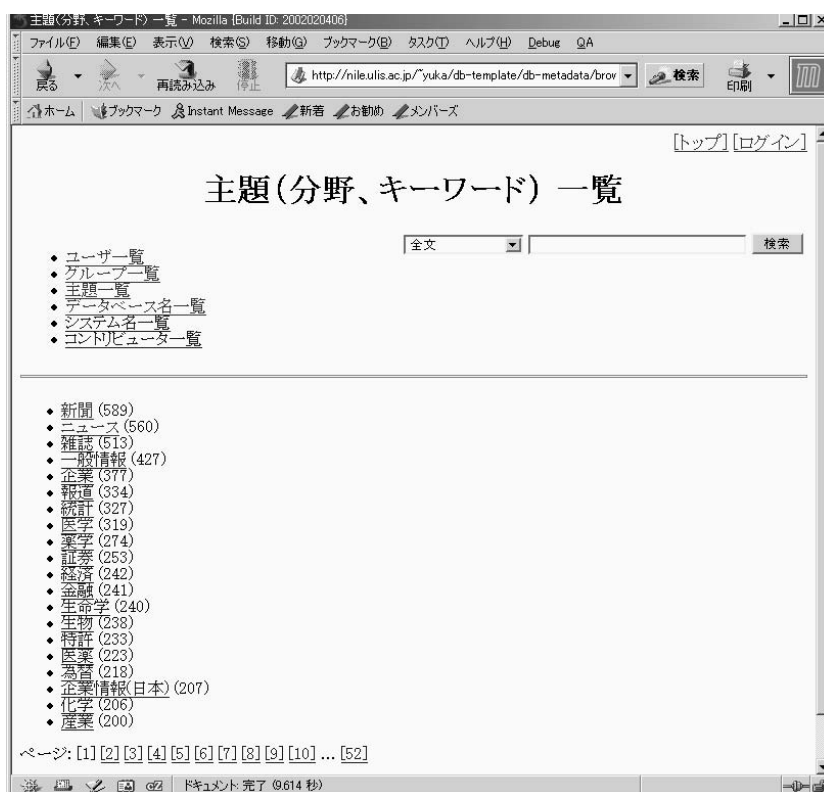


図 4.6 主題一覧

図 4.6 は、主題一覧の画面である。全データベースの基本情報に出てくる主題の頻度の多いものから順に 20 件ずつ表示しており、各主題とその出現回数を表示している。画面下部の数字のリンクを辿ると他の 20 件が表示される。たとえば、「新聞」という主題は全基本情報中で 589 回出現している。「新聞」のリンクを辿ると「新聞」という主題を持つデータベースが簡易表示される。

ユーザページ表示

ユーザページ表示では、ユーザ毎に作成した基本情報や後の 4.4.6 節で説明するグループ化機能にて作成したグループ情報の一覧を表示する。また、ログインした状態であれば、ユーザ自身のユーザページへの移動は、各画面のトップに、ユーザ自身のページへのリンクがあるので、そのリンクを辿ることですぐさま移動することができる。

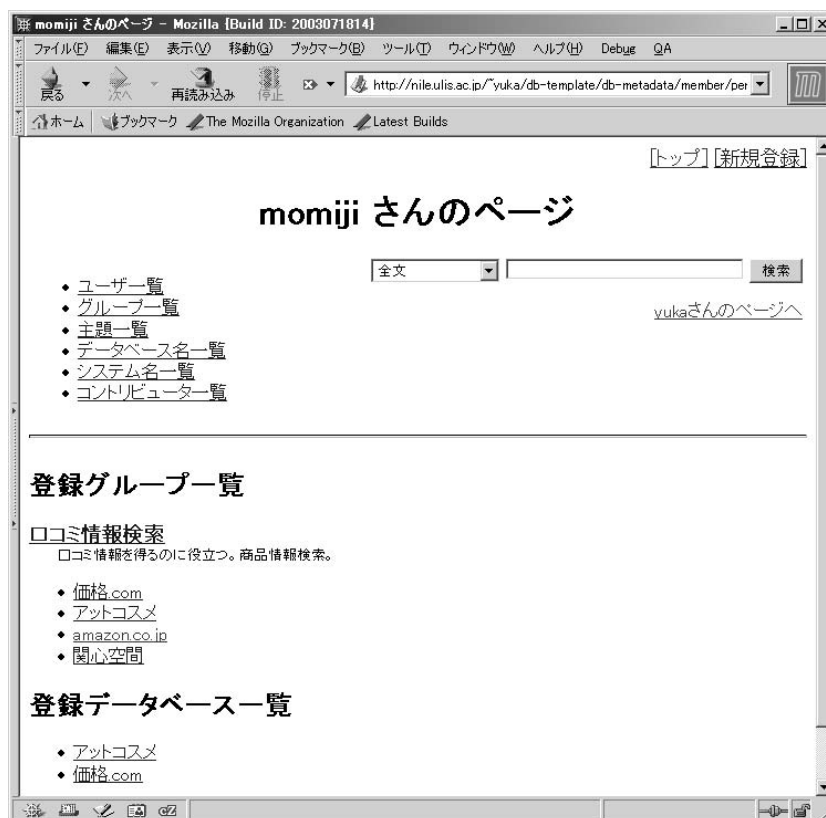


図 4.7 ユーザのページ

図 4.7 はユーザ momiji のページである。ユーザ momiji の登録したグループは「口コミ情報検索」であり、このグループに属するデータベースは「amazon.co.jp」など 4 つあることがわかる。また、登録した基本情報は 2 つあり、「アットコスメ」と「価格.com」があることがわかる。

4.4.3 検索機能

検索機能は、システムに入力された基本情報を検索する機能であり、全項目について全文検索するかデータベース名、主題、システム名、コントリビュータ名の項目中から一つのフィールドを指

定して検索するかを選択して、検索できる。フィールド指定の検索を使うことにより、ユーザがあらかじめデータベース名を知っている場合や、同じシステムによるデータベースを知りたい場合のアクセスが容易になる。

4.4.4 基本情報の登録・更新機能

ユーザによる基本情報の登録・更新機能には、新規登録、更新登録、取り込み登録(流用入力)の3種類の登録方法がある。まず、新規登録は、基本情報を一から入力し、新規に登録する機能である。更新登録とは、既にシステムに入力されている基本情報のうち、変更のある部分だけを編集して、更新して登録する機能である。取り込み登録(流用入力)とは、既にシステムに入力されている基本情報を取り込み、必要な部分だけ編集して、異なる基本情報として、新規に登録する機能である。取り込み登録の場合は、情報源を明らかにするために、取り込んだ元の基本情報のIDを記録しており、詳細情報表示のときに、取り込み元の基本情報にすぐに辿れるようになっている。

図 4.5 は「取り込み入力」で作成した基本情報の詳細表示である。「この情報は 0632 を元に作成しています」の 0632 のリンクを辿ると、取り込んだ元のデータベースの基本情報の詳細表示をみることができる。

4.4.5 コメント機能

コメント機能は、各データベースの基本情報にコメントを登録する機能である。これはいわゆる電子掲示板機能であり、ユーザは、この機能を使って、他のユーザなどと話し合ったり議論したりしながら、データベース利用の際のちょっとしたコツなどを共有していくことができる。

図 4.8 は、投稿の入力フォームに「検索結果の行頭の 001 などの数字はなにをあらわしているのですか?」を入力しているところである。図 4.9 は、「投稿ボタン」をクリックしてコメントを投稿して表示されたところである。

4.4.6 グループ化機能

グループ化機能は、複数のデータベースを関連付け、グルーピングする機能である。グループの情報は、以下の構成要素から成り立っている。

- グループ名
- グループの説明
- グループに属するデータベースの集合
- グループを作成したユーザ ID

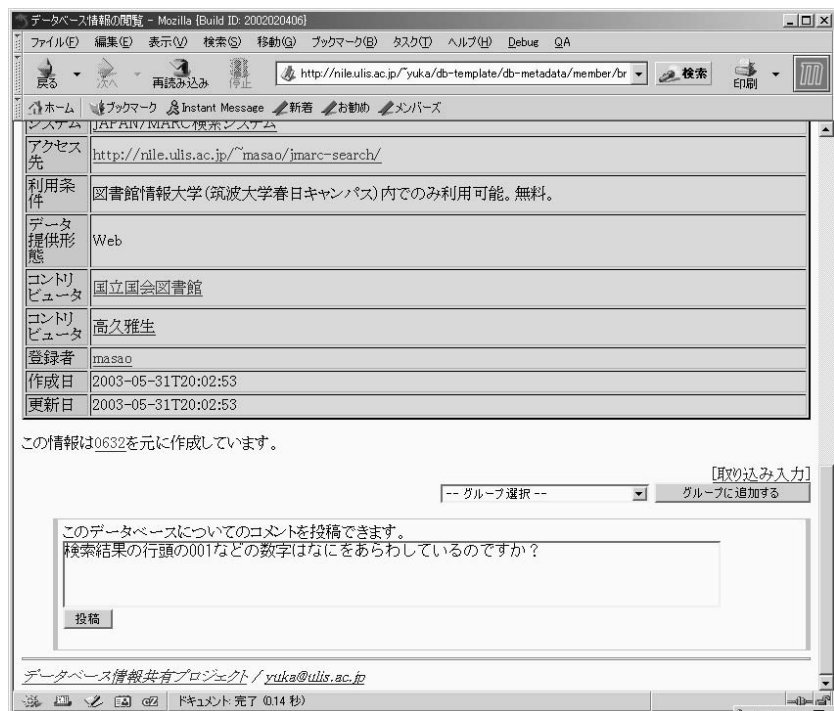


図 4.8 基本情報詳細表示における操作例 (1)

ログインユーザは、各自がグループを作成し、データベースをグルーピングすることができる。グルーピングされたデータベースは、各データベースの詳細表示から迎れる。

この機能を利用することにより、ユーザは自分なりの観点でデータベースをまとめ、分類し整理することができる。この整理した情報を共有することができる。

新規にグループを追加したり、グループの名前やグループの説明を変更する場合はユーザ自身のページで行なう。

図 4.10 は、ログインユーザ yuka の ユーザ自身のページにて、新規グループ名に「情報学関連」、グループの説明に「情報学関連で有用なデータベース」を入力しているところである。ここで「登録」ボタンをクリックするところである。ここで、「グループに追加する」ボタンをクリックすると、図 4.11 のように、表示され、先程作成したグループができていのがわかる。ここで、「修正」リンクを辿ると、グループ編集画面に移る。

図 4.12 は、ログインユーザ yuka の ユーザ自身のページにて、グループ名:「お気に入り」の「修正」リンクを辿ってグループ編集画面を表示したところである。ここで、グループ名やグループの説明の修正を各入力フォームで行なって、「変更」ボタンをクリックすると、グループの修正が行なえる。「グループの削除」をクリックすると、このグループを削除する。それぞれボタンをクリックすると、ユーザ自身のページに戻る。

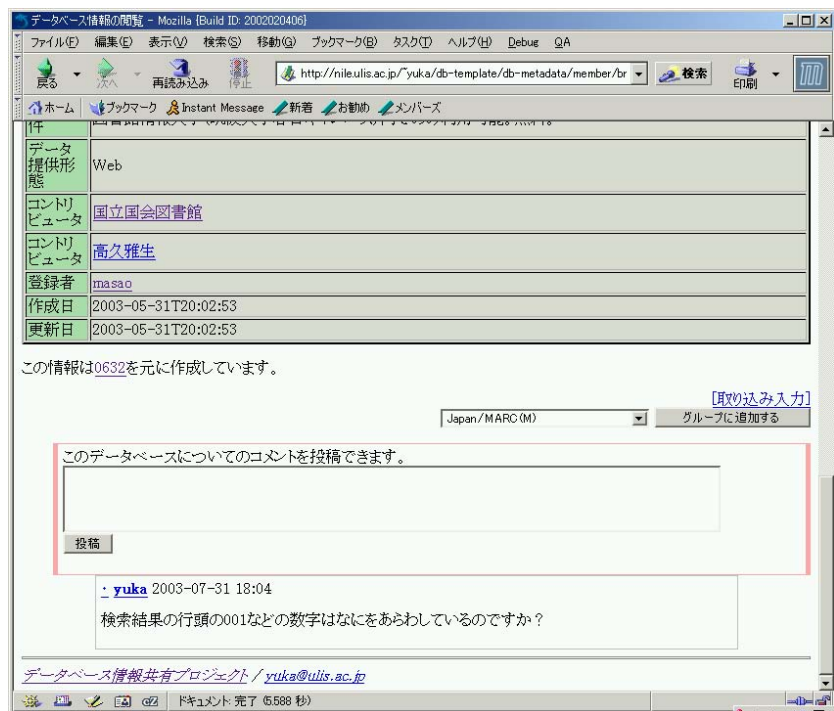


図 4.9 基本情報詳細表示における操作例 (2)

下部の数字のリストの数字のリンクはデータベースの ID を表わしており、このリンクを辿るとその ID のデータベースの基本情報詳細表示画面に移り、このグループに属している基本情報を見ることができる。このグループから特定の基本情報を外したいときは、数字の横の「削除」のリンクを辿ればよい。

図 4.9 は、追加するグループ「Japan/MARC(M)」を選択しているところである。ここで、「グループに追加する」ボタンをクリックすると、図 4.13 のように表示される。「yuka」はグループを作成したユーザ ID、「Japan/MARC(M)」はグループ、「データが Japan/MARC(M) と思われるもの (データ提供形態やシステムは異なる)」はグループの説明を表わす。ユーザ ID はリンクになっており、ユーザ ID 「yuka」を辿るとそのユーザのユーザページ表示を閲覧でき、このグループを作成したユーザが他にどのようなグループを作成しているか、どのような基本情報を登録しているかがわかるようになっている。また、グループもリンクになっており、グループ「Japan/MARC(M)」を辿ると、そのグループに属するデータベースの簡易表示を閲覧できる。

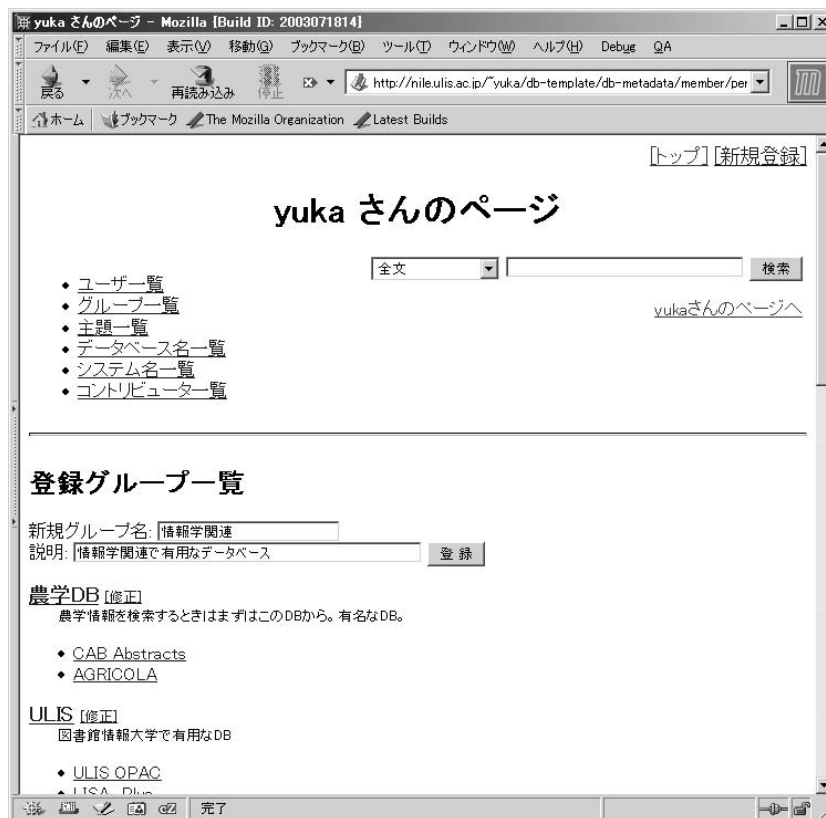


図 4.10 ユーザ自身のページ (新規グループ作成)

4.4.7 ユーザ認証

ユーザ認証では、ログイン ID とパスワードを使ってシステム内に予め登録されたユーザを識別する機能である。本システムでは登録していないユーザはゲストユーザとなり、ログイン ID を持っているのはログインユーザとなる。基本情報の登録、グループ化、コメント機能などは、全て、このログインユーザのみが行なえる機能である。また、基本情報やグループ化情報は、各個人毎に、自身が登録したもののみを変更できることとし、他のゲストユーザやログインユーザは閲覧のみとし編集できないようにした。また、各ユーザが自分が変更可能な権限がある場合のみ、編集画面へのリンクが出るなど、権限の有無がユーザインタフェースにも反映されるようにした。

例えば、図 4.5 での詳細表示では、ログインしていない状態での表示のため、グループへの追加やコメントの追加はできないが、図 4.8 では、ログインした状態のため、これらの入力が可能になっている。他の例を挙げると、図 4.7 では、ユーザ yuka でログインしており、ユーザ momiji のグループ新規作成権限はないため、グループの新規作成のためのフォームが表示されていないが、

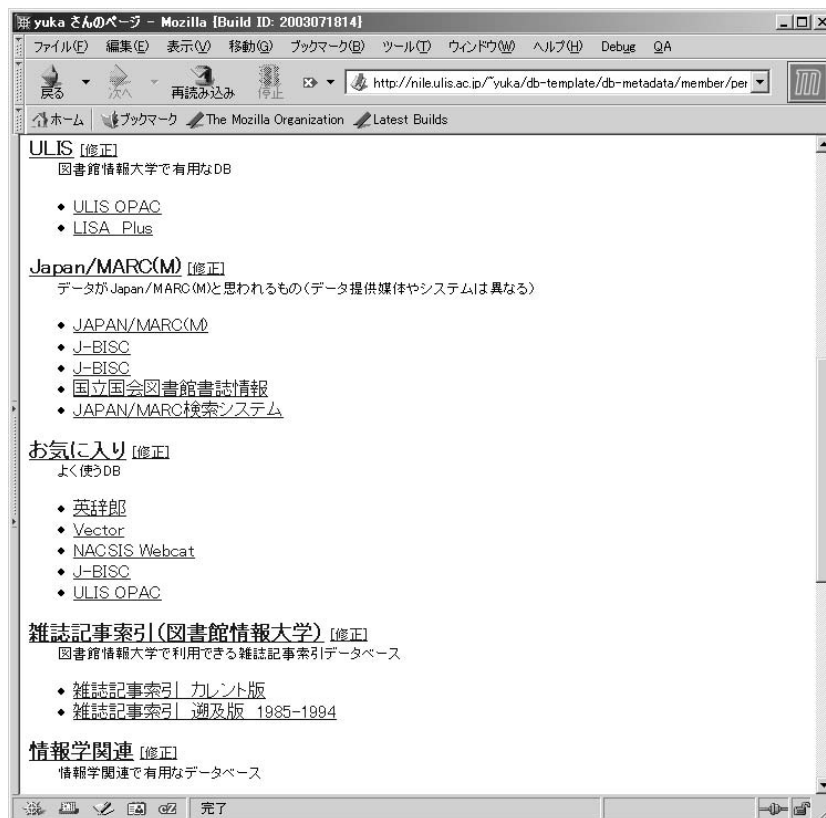


図 4.11 ユーザ自身のページ (新規グループ作成後)

図 4.10 では、グループ新規作成のためのフォームが表示されているところである。

4.5 考察

本節では、まず本研究と類似した手法をとっている既存の研究やシステムについて、その類似点・相違点を述べ、次に本システムの運用方法について述べる。

4.5.1 情報共有システム

本研究ではデータベースに関してその基本情報・付加情報を設定し、それらの情報をデータベース開発者・一般利用者双方で共有できる枠組を考案し、システム構築を行った。このように開発者・一般利用者双方で情報共有を行って双方の利便を高める枠組みの提案は、これまでデータベースに関する情報の分野では見当たらなかった。データベース以外の別分野における、このような情報共有の取組みには、ソフトウェア開発の分野における Sourceforge.net[88] や Freshmeat.net[89]

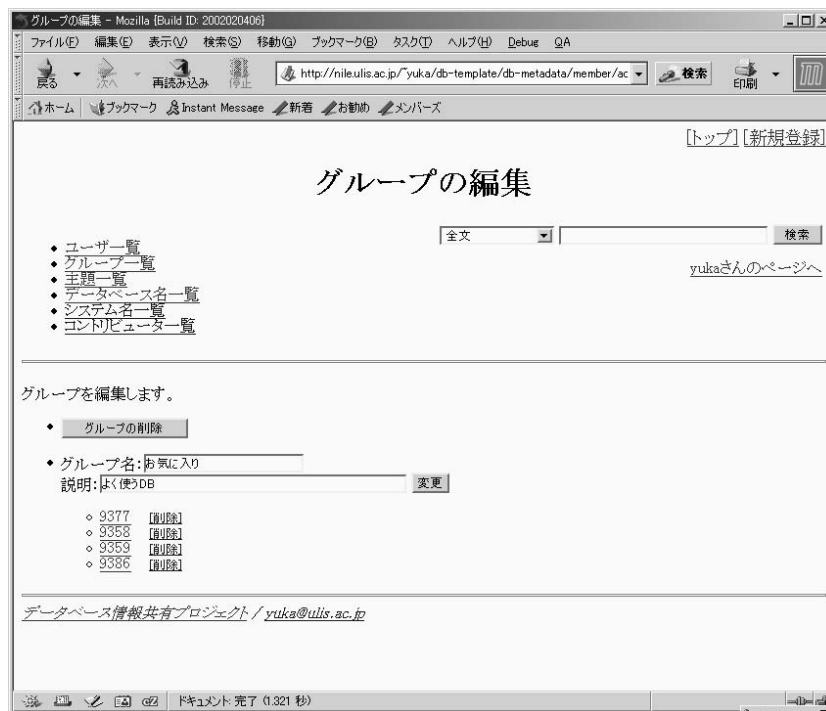


図 4.12 グループ編集

によるものがある。

Sourceforge は、オープンソースソフトウェアの効率的な開発のために提供されている一種のポータルサイトであり、ソフトウェア開発者に対して WWW/FTP/ML/CVS/BBS といった機能を提供することにより、ソフトウェアのダウンロードや開発のための議論がおこなえる基盤を提供している。また、Freshmeat はオープンソースソフトウェアに関する情報を共有するサイトであり、主にソフトウェアのリリース情報の共有を行っている。Sourceforge/Freshmeat の両サイトとも、各ソフトウェアの情報については、名称、概要、開発状況（アルファ・ベータ・安定など）、使用のためのライセンス条件、使用言語、動作環境、開発言語、カテゴリといった項目が設けられている。2004 年 9 月現在、Sourceforge 上では約 16,000 件、Freshmeat では約 34,000 万件もの多くのソフトウェアが登録されており、開発者・利用者ともに多くが参加するポータルサイトとなっているため、開発者にとっては他の開発者・一般利用者の双方とコミュニケーションを取りながら開発を進められる。一方、一般利用者にとっても自身が使用したいソフトウェアを探索することができる。

その他の分野を対象とした情報共有手法には、基本情報が入力されており、利用者がコメントなどを入力するコミュニティサイトがある。代表的なものには、アットコスメ [90]、価格.com [91]、amazon.co.jp [92] などがある。これらコミュニティサイトは、本研究における付加情報の共有手法と似ているが、基本情報の入力にはできない。また、アットコスメと価格.com は基本情報の再利用

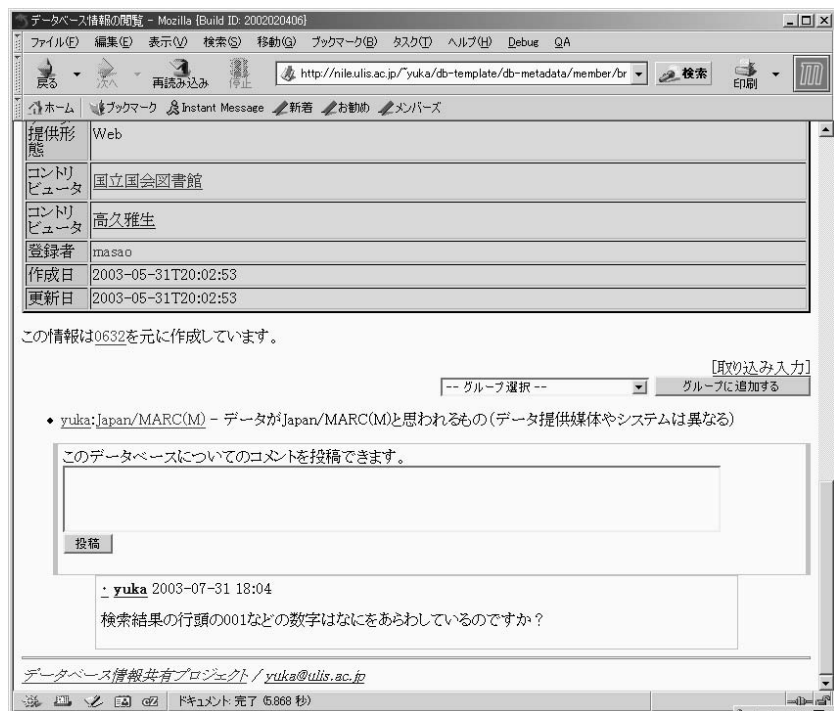


図 4.13 基本情報詳細表示における操作例 (3))

もできない。

一方、より自由な形式でユーザ同士によるコミュニティ構築が行なえるシステムとして Wiki[93] がある。Wiki は Wiki サイトと呼ばれる共同作業用サーバを用意することでそのサイトの訪問者が自由に書き込んだりページを新たに作成したりできる特徴を持ち、ページ間のハイパーリンクを容易に作成できるようにしたものである。また、関心空間 [94] は、利用者同士が自分の興味のあるキーワードを登録し、さらに、キーワード同士につながりをつけられるシステムである。これらのシステムはユーザの様々な観点に基づく情報共有が行なわれているという点で、本研究における付加情報の共有手法と共通点があるが、これらには本システムのような基本情報を再利用する機能はない。

4.5.2 基本情報の共同入力・共同利用

本研究のように、基本情報を共同で入力することにより構築されているシステムとして OCLC[95] や NACSIS[96] による共同目録作業システムがある。これらは書籍の目録を様々な図書館で共同で入力することにより、それを共有して利用できるようにしている。このシステムを利用して、所蔵図書館を検索することで欲しい本を相互貸借することもできるし、既にどこかの図

書館で目録情報（本研究での基本情報にあたる）が入力されていれば、他の図書館は基本情報を取り込むだけで、図書館ごとに所蔵図書の基本情報を入力しなくてもすむ。このように基本情報の共同入力・共同利用のアプローチは書籍を対象として広く行なわれている。

本研究では、このアプローチをデータベースの基本情報にも適用した。この場合、本は一冊という物理的に固定的な対象のためいったん基本情報を入力すれば変更することはないが、データベースは本とは異なり、データは常に更新される、システムが変更される、データベースのサービスが停止するなど流動的である。また、データベースは様々な形態でサービスされるため、データベースを同定することは難しい。そこで、これまでの書籍での仕組みをそのまま適用できるわけではない。

そこで本システムでは、システムの側ではデータベースの同定は行なわないこととし、一つのデータベースに対して、複数の基本情報を入力可能とした。また、同じような基本情報が増えてしまう問題に対しては、グループ化機能を利用して同じと思われるデータベースをひとくくりにしたり、取り込み入力で元にした基本情報がある場合には、取り込み元の基本情報がわかるようにするなどの機能により、対応できるようにした。

4.5.3 本システムの運用方法

情報を共同入力・共同利用する場合、登録可能な人に一定の制限を加えたり、登録の際にルールを設けるなど、様々な運用方法が考えられる。どのような運用方法がよいと一律に規定できるものではなく、システム運用主体の方針により判断されるべきものである。そこで本研究では、様々な運用方法に対応するシステムを構築した。本節では、本システムの運用方法における事項について述べる。まず、システム運用に際して考慮すべき事柄について述べ、次に、様々な運用方法における本システムの対応可能性について述べる。

本システムの運用方法は、情報の作成と利用の2つの観点から見ることができる。まず、情報の作成においては、情報の質と量は概してトレードオフの関係にあり、情報の質や量をコントロールするための要素には、情報登録者の制限、情報作成ガイドラインの作成、登録情報のチェック、統制語の作成などがある。本研究では、これらの要素の組み合わせ全てについての比較検討は行なわないが、これらを組みあわせて、レベルを設定することで、質や量がコントロールでき、よりよい基本情報の共有ができるものとする。また、情報の利用という点からは、後述するように利用者の制限を設けることが挙げられる。

次に、上述した運用方法の要素の本システムにおける対応可能性について述べる。まず、情報登録者の限定を考えた場合、例えば、質を上げことを期待して、基本情報の登録者をデータベースに詳しい人に限定する。または、量を増やすことを期待して、登録者の間口を広げ、誰でも登録できるようにするなどが考えられる。

本システムでは、ログインユーザは誰でも基本情報を登録・更新できるようにし、ログイン ID

を持っていないゲストユーザの場合でも、閲覧・検索は利用可能にすることで、システム変更なしに、システムの運用方法を変えるのみで登録者を制限可能にした。例えば登録者をデータベースに詳しい人に限定したいなどという場合でも、ログインユーザを限定するという運用方法をとることで実現できる。

情報作成ガイドラインの作成については、例えば、基本情報を作成するための情報作成ガイドラインであれば、対象とするデータベースの定義、各項目の具体例を挙げるなどがある。この情報作成ガイドラインを本システムにおいて対応させる場合は、情報入力の際にヘルプなどとして、ガイドラインを提示するという方法が考えられるが、本システムではこのようなガイドラインとの連携機能には対応していない。

登録情報のチェックとは、例えば、登録された基本情報が正しいかどうかをチェックする人を設け、指摘、修正することなどを指す。本システムにおいてこれに対応させるには、チェックした人がコメント機能を使って間違いの内容を指摘することができる。本システムでは、情報登録者のみが情報を修正できるため、チェックした人が修正することはできない。このため、登録情報チェックの作業者に、他人の登録情報を修正する特別な権限を持たせるには、システムの変更が必要である。

また、利用者全体の限定を考えた場合、例えば、使っているデータベースや、使い方が外部にもれることが困る企業の場合、利用者全体を企業内のみに限定する。または、データベース提供者が自身の提供しているデータベースを広報して、多くの人にデータベースを使ってもらいたいため、誰でも利用可能にする。などが考えられる。本システムは、WWW の CGI で実現されているため、WWW サーバの閲覧可能な範囲の設定を変更するのみで、本システムを変更することなく対応可能である。

このように、データベースに関する情報の共有は、本システムの運用の仕方により、様々なレベルで行なうことができる。

また、いろんな運用方法がとれることにより、様々な分野や用途にこのシステムを適用することが可能となる。以下にその例を挙げる。

- 大学の附属図書館がシステムを運営し、自館の利用者のサービス向上のためにシステムを利用する
- ある主題分野のコミュニティがシステムを運営し、同じ主題分野の研究者同士で情報を共有するためにシステムを利用する
- 多くのデータベースを提供しているデータベースプロバイダがシステムを運営し、自社が提供しているデータベースの宣伝、利用サポートのためにシステムを利用する
- データベースを提供している機関とは別に、利用者側である第3者がシステムを運営し、無償で利用可能なデータベースについての情報を不特定多数の利用者間で共有する

4.6 まとめ

データベースの基本情報を設計し、その基本情報を入力・閲覧できるシステムを作成した。また、基本情報だけでなく、ユーザ毎のノウハウなど入力・閲覧できるシステムを作成した。このことによりデータベースに関する情報の共有が可能になり、検索要求に適したデータベースの発見・利用が容易にできるようになった。

第 5 章

結論

本研究では、多種多様なデータベースが分散的に存在し、データベースの利用者が多様になったゆえにおこる問題の中でも、以下の 2 つに着目し、

1. 利用者のニーズに合ったデータベースを探し出し選択することの難しさ
2. データベースを使いこなすことの難しさ

これらの問題を、以下の 2 つのアプローチで解決できると考え、

- 検索ユーザインタフェースの統一
- データベースに関する情報の共有

大きくわけて、以下の 3 つの研究を行ない、このアプローチの有効性を示した。

第 1 の研究は、Z39.50 検索クライアントの構築である。情報検索の相互運用を考慮した通信プロトコル Z39.50 に対応し、日本語を使って検索し、検索結果を表示でき、ネイティブに動く Z39.50 検索クライアントを開発することにより、利用者が統一的なユーザインタフェースで、複数のデータベースを検索できることを示した。

第 2 の研究は、Z39.50 データベース選択支援環境である。この研究は WWW と Z39.50 を透過的に利用できるデータベース選択支援環境を提案し、その環境を実現する WWW-Z39.50 クライアントを開発した。この環境では Z39.50 データベースに関する情報を WWW 上で共有して利用できるだけでなく、利用者が自由にデータベースを組織化し、利用者同士でお互いに参照・利用できる。また、共有された情報であるデータベースリストの中から任意の一つをクリックすることによって、該当するデータベースの検索画面が起動し、直ちに検索を開始できる。

第 3 の研究は、データベースに関する情報を利用者同士で共同入力・共同利用できるシステムを開発した。この研究では、まず、データベース選択に必要な情報を基本情報、付加情報の 2 つに分け、21 項目からなる基本情報を設計するとともに、この基本情報を入力・閲覧できるシステムを作成した。さらに、複数のデータベースをさまざまな観点でグループ化した情報やデータベース毎

の利用方法などからなる付加情報についても、入力・閲覧できるようにした。また、データベース台帳総覧のデータ 3,614 件に適用し、その有効性を示した。

このように本研究では、分散的に存在する多種多様なデータベースを多様な利用者が利用する状況において、検索ユーザインタフェースの統一とデータベースに関する情報の共有というアプローチで研究してきた。

検索ユーザインタフェースを統一することによって、データベース毎にユーザインタフェースが変わることなく、利用者は複数のデータベースを同様に利用できる。このことは個々の利用者にとってよいだけでなく、データベースを提供する側にもよい。それは、今迄、新たなユーザインタフェースの習熟ができないために、利用しなかった利用者が利用できるようになるため、自身の提供したデータベースを使ってもらえる機会が増えるからである。本研究では、第 1 の研究で、検索ユーザインタフェースの統一を実現し、その有効性を示した。

データベースに関する情報を共有することによって、日々変化するデータベースについての情報を個人や組織が個々に追うことなく、情報を共有している人々全員がそれぞれ得意分野を担当するなどして、相互に補完することができることは、利用者にとって有用である。また、多くのデータベースを提供することは図書館などでは普通になってきており、図書館の利用者のために、データベースに関する情報を個々の図書館が提供しなければならなくなっている。そのため、データベースに関する情報を作成するコストが大きくなったり、同じデータベースに関する情報を重複して作成するような状況になっている。そこで、図書館のようなデータベースに提供する側でも、個々の提供機関におけるデータベース情報作成のコストを低減し、効率よく情報を提供する点において、データベースに関する情報を共有する手法は有用である。本研究では、第 2、第 3 の研究において、データベースに関する情報の共有を実現し、その有効性を示した。

検索ユーザインタフェースの統一とデータベースに関する情報の共有のそれぞれが可能になるだけでなく、これら 2 つを組みあわせれば、より効果的に多種多様なデータベースの利用支援になる。つまり、片方だけが可能であっても、検索できないことがあるからである。検索ユーザインタフェースの統一がされていれば、利用者はどのデータベースにも同様の使いかたで検索できるが、データベースの内容がわからなければ、効率的な検索が行えないし、また、逆にデータベースに関する情報の共有によって、データベースの内容がわかっているにもかかわらず、データベース毎に使い方を覚えなければ、検索ができない。例えば、そのデータベースが厳密な統制語で制御されているデータベースの場合、フリーキーワードで検索して、ヒットする結果が出ないことがある。この問題を解決するには、データベースの内容（データベースが厳密な統制語で制御されている）を知らなければ、検索できないなどである。本研究では、検索ユーザインタフェースとデータベースに関する情報の共有の両方を第 2 の研究において実現し、その有効性を示した。

現在、本研究を始めた当時より、ますます、Web 上における情報流通は盛んになり、また、多くの有用なデータベースが手軽に利用できるようになっている。

Web 上における情報流通の顕著な例としては、Weblog[97] がある。これは、個人の情報発信環境をより手軽に展開でき、RSS のようなメタデータ技術やトラックバック [98] といったコミュニケーション技術の進展とともに発展を遂げている。Weblog 自体は、個人が持つ情報の発信を簡便に行うことを目的として、開発が進められたため、雑多な情報をまとめて発信し、コミュニティ内で情報を交換したり共有したりできることに主眼が置かれ、本研究のようにデータベース情報に特化していないため、データベース情報の共有、交換に必要な基本情報・付加情報といった観点からの研究は見当たらない。しかし、Weblog において開発されてきた複数の技術のうちには、情報共有という観点から、本研究と関連するものもある。例えば、Weblog によるコンテンツは、RSS メタデータをソフトウェアを用いて複数のサイトから自動的に収集し、一覧することが可能になる。RSS メタデータ形式は、Weblog ツールに実装されたことにより普及したといわれている。同様に、本研究で開発したデータベースに関する一種のメタデータである基本情報・付加情報も本研究で開発したような利用者間での共有システムで扱うことができるようになり、普及が進めば、RSS と同様にソフトウェアを用いて自動的に収集し、一覧するなどの活用が可能となるだろう。このように、Weblog が持つ情報共有に基づく情報発信機能という考え方は、本研究で開発した情報共有システムにも応用が可能であるだろう。

多くの有用なデータベースが手軽に利用できるようになった例としては、例えば、筑波大学の付属図書館が筑波大学内に提供しているデータベース数は、100 を越している。また、雑誌・論文が電子的に提供されることも急激に増えており、データベースを検索して見つけた論文の全文が、図書館に行かずとも、読めるようにもなっている。このように、分散環境におけるデータベースの利用支援は、ますます重要になってきており、本研究の意義は大きい。

参考文献

- [1] 田澤 由利. “北見工業大学『情報科学概論』実習課題 回答”, 5012.jp, (オンライン), 入手先 <http://www.5012.jp/search/search_2.htm>, (参照 2003-12-08).
- [2] Bergman, Michael K. The Deep Web: Surfacing Hidden Value. The Journal of Electronic Publishing. Vol.7, No.1, 2001, (ISSN 1080-2711), (online), available from <<http://www.press.umich.edu/jep/07-01/bergman.html>>, (accessed 2003-12-08).
- [3] Date, C. J. An introduction to database systems. 5th ed. Addison-Wesley, 1990, 854p. (ISBN 0-201-51381-1).
- [4] Garcia-Molina, Hector; Ullman, Jeffrey D.; Widom, Jennifer. Database systems: the complete book. Upper Saddle River, Prentice Hall, 2002, 1119p. (ISBN 0-13-031995-3).
- [5] データベース振興センター編. データベース白書 2003. 東京, データベース振興センター, 2003, 413p. (ISBN 4-924777-17-X).
- [6] 国立国会図書館. “対象とするデータベース”, 国立国会図書館データベース・ナビゲーション・サービス Dnavi, (オンライン), 入手先 <http://dnavi.ndl.go.jp/Dnavi_Intro.html#INTRO_3>, (参照 2004-09-04).
- [7] 情報検索(Z39.50)応用サービス定義及びプロトコル仕様 JIS X 0806:1999 (ISO 23950:1998). 東京, 日本規格協会, 1999, 272p.
- [8] Apple Computer. Human interface guidelines : the apple desktop interface. イントランス訳. 東京, アジソンウェスレイパブリッシャーズジャパン, 1989, 147p. (ISBN 4810180182).
- [9] Microsoft Corporation. Windows ユーザーインターフェイスデザインガイド. 森川創訳. 東京, アスキー, 1995, 523p. (ISBN 4756116140).
- [10] 宇陀 則彦, 江草 由佳, 高久 雅生, 石塚 英弘. Z39.50 による日本語書誌データ検索システム. 情報知識学会誌. Vol.9, No.2, 1999, p.1-15.
- [11] 江草 由佳, 高久 雅生, 宇陀 則彦, 石塚 英弘. Z39.50 データベース選択支援環境. 情報知識学会誌. Vol.11, No.2, 2001, p.1-10.
- [12] EGUSA, Yuka; TAKAKU, Masao; ISHIZUKA, Hidehiro. Network Community Oriented Information Sharing System for Databases. Proceedings of International Symposium on Digital Li-

- braries and Knowledge Communities in Networked Information Society DLKC'04. Tsukuba, 2004-03, University of Tsukuba, p.64-71. (online), <<http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp64.pdf>>, (accessed 2004-09-19).
- [13] 江草 由佳, 高久 雅生, 石塚 英弘. 分散環境におけるデータベースに関する情報の共有システム. 情報知識学会誌. Vol.14, No.3, 2004, p.1-17.
- [14] Z39.50 Maintenance Agency. ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50) : Application Service Definition and Protocol Specification. 1995, 156p. (online), available from <<http://www.loc.gov/z3950/agency/1995doce.html>>.
- [15] Lynch, Clifford A. The Z39.50 Information Retrieval Standard : Part I: A Strategic View of Its Past, Present and Future. D-Lib Magazine. 1997. (online), available from <<http://www.dlib.org/dlib/april97/04lynch.html>>, (accessed 2004-09-19).
- [16] 上田 修一. Z39.50 とその可能性. 情報の科学と技術. Vol.48, No.3, 1998, p.126-133.
- [17] 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, 1994, p.4.
- [18] 上田 修一. Z39.50 の可能性と問題点. 三田図書館・情報学会 1996 年度研究大会予稿集. 1996, p.37-40. (オンライン), 入手先 <<http://www.slis.keio.ac.jp/~ueda/z3950/mita96.html>>, (参照 2000-01-13).
- [19] 松林 正己. 標準情報検索プロトコル Z39.50 の国際的展望. 情報の科学と技術. Vol.48, No.3, 1998, p.144-155.
- [20] Index Data. “Z39.50 Target Information”. updated 1999-06-30, (online), available from <<http://www.indexdata.dk/targettest/targetstat.shtml>>.
- [21] 牧野 順子. Z39.50 対応電子図書館パッケージ・システム : OCLC SiteSearch System. 情報の科学と技術. Vol.48, No.3, 1998, p.172-174.
- [22] 佐藤 康之. MARUZEN Z39.50 Gateway—CAT—Z39.50 商用サービス利用事情—. 情報の科学と技術. Vol.48, No.3, 1998, p.167-171.
- [23] 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, 1997, p.692-701.
- [24] 図書館情報大学附属図書館. 図書館情報大学デジタル図書館 (ULIS-DL). (オンライン), 入手先 <<http://lib.ulis.ac.jp/>>, (参照 1999-05-08).
- [25] Waseda University Library. 早稲田大学学術情報システム (WINE). (オンライン), 入手先 <<http://wine.wul.waseda.ac.jp/>>, (参照 1999-05-08).
- [26] 東京工業大学附属図書館. Titech Digital Library 東京工業大学電子図書館. 更新 1999-03-25, (オンライン), 入手先 <<http://tdl.libra.titech.ac.jp/>>.
- [27] 小島 明, 田中 純, 篠原 正紀, 落合 崇道. 大学での取り組みが進む電子図書館. NTT 技術ジャーナル. Vol.11, No.1, 1999, p.55-58.
- [28] 安達 淳. 電子図書館成立の条件. デジタル図書館. No.1, 1994, p.23-27. (オンライン), 入

- 手先 <http://www.dl.ulis.ac.jp/DLjournal/No_1/adachi/adachi.html>,
(参照 1998-07-23) .
- [29] 学術情報センター事業部データベース課. 電子図書館サービスの実施状況. 学術情報センター
ニュース. No.47, 1998, p.9.
- [30] 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, 1998, p.222-228.
- [31] 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.
- [32] 石田 茂. Z39.50 と日本語書誌目録の連携に関する考察. 情報処理学会研究報告 99-DBS-
118//99-FI-54. Vol.99, No.39, 1999, p.81-88.
- [33] 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
研究会報告 99-FI-44. Vol.96, No.116, 1996, p.9-16.
- [34] 江草 由佳, 真野 泰久, 宇陀 則彦, 石塚 英弘. 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. 第 6 回研究報告会講演論文集. 情報知識学会. 東京, 1998-05. 情報知識学会, 1998,
p.29-36.
- [35] プロトコルハンドブック編集委員会. 新プロトコルハンドブック. 東京, アトソン, 1994,
241-251p.
- [36] 安齋 宏幸. Z39.50 の技術解説. 情報の科学と技術. Vol.48, No.3, 1998, p.134-139.
- [37] Index Data. “The YAZ Toolkit”. updated 1999-03-22, (online), available from <<http://www.indexdata.dk/yaz/>>.
- [38] 高林 哲. 全文検索システム Namazu. 更新 1999-03-15, (オンライン), 入手先 <<http://openlab.ring.gr.jp/namazu/>>.
- [39] 馬場 肇. 日本語全文検索システムの構築と活用. 東京, ソフトバンク, 1998, 258p. (ISBN
4-7973-0691-2) .
- [40] 安齋 宏幸. WWW による JAPAN/MARC の提供実験. つくば, 図書館情報大学, 1994. 卒業論文.
- [41] 黒澤 正彦, 西村 徹編. マークをうまく使うには: 機械可読目録入門. 東京, 三洋出版貿易, 1986,
310p. (ISBN 4-87930-030-6) .
- [42] 国立国会図書館. Japan/MARC マニュアル: 図書編. 第 2 版. 東京, 国立国会図書館, 1998, 186p.
(ISBN 4-87582-527-7) .
- [43] 奈良先端科学技術大学院大学自然言語処理学講座. 日本語形態素解析ツール・茶筌. (オンライ
ン), 入手先 <<http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>>, (参
照 1999-05-08) .
- [44] Index Data. “YAZ User’s Guide and Reference: Introduction”. (online), available from <<http://www.indexdata.dk/yaz/yaz.shtml>>, (accessed 1999-05-04) .

- [45] 安斎 宏幸. インターネット環境における日本語書誌情報システムの構築. つくば, 図書館情報大学, 1997. 修士論文.
- [46] 真野 泰久. Z39.50 プロトコルを用いた検索サーバの開発. つくば, 図書館情報大学, 1998, 92p. 卒業論文.
- [47] Hasebe, Kigen; Nakamoto, Ken'iti; Yamamoto, Takeo. An information retrieval system on internet for languages without obvious word delimiters. Proceedings of International Symposium on Digital Libraries 1995. Tsukuba, 1995-08, University of Library and Information Science, 1995-08. p.181-185.
- [48] Library of Congress. "Character Set and Language Negotiation (2)". Z39.50 Maintenance Agency. 1998-11, (online), available from <<http://lcweb.loc.gov/z3950/agency/defs/charsets.html>>.
- [49] Denenberg, R.; Kunze, J.; Lynch, D. eds. "RFC 2056 : Uniform Resource Locators for Z39.50". The Internet Engineering Task Force. 1996-11, (online), available from <<http://www.ietf.org/rfc/rfc2056.txt>>.
- [50] 九州大学. 多機関 OPAC 横断検索.(オンライン), 入手先 <<http://zeus.lib.kyushu-u.ac.jp/htdocs/Qindex.html>>, (参照 1999-05-08).
- [51] Library of Congress. "About profiles". Z39.50 Maintenance Agency. (online), available from <<http://www.loc.gov/z3950/agency/profiles/about.html>>, (accessed 2004-09-06).
- [52] The Bath Profile Editorial Team; Lunau, Carol; Miller, Paul; E. Moen., William. The Bath Profile, (online), available from <<http://www.collectionscanada.ca/bath/bp-current.htm>>, (accessed 2004-09-04).
- [53] 千原 秀昭, 時実 象一. 化学情報—文献とデータへのアクセス—. 東京, 東京化学同人, 1991, 216p. (ISBN 4-8079-0357-8).
- [54] Shulz, Hedda; Georgy, Ursula. ケミカル・アブストラクツと CAS ONLINE の活用法. 時実 象一訳. 東京, 日外アソシエーツ, 1995, 280p. (ISBN 4-8169-1329-7).
- [55] 笹本 光雄. 医薬品情報. 東京, 広川書店, 1981, 305p.
- [56] 尾城 孝一. 東京工業大学電子図書館 (TDL: Titech Digital Library). デジタル図書館. No.16, 1999, p.24-38. (オンライン), 入手先 <http://www.dl.ulis.ac.jp/DLjournal/No_16/3-ojiro/3-ojiro.html>, (参照 2004-09-19).
- [57] 情報サイエンス研究会. インターネットと電子メールによる医学・薬学情報収集活用ガイド. 東京, ソシム, 1998, 271p. (ISBN 4-88337-059-3).
- [58] Cornell University Library. Cornell University Library Gateway. (online), available from <<http://campusgw.library.cornell.edu/>>, (accessed 2000-01-09).
- [59] Payette, Sandra D.; Rieger, Oya Y. Z39.50: The User's Perspective. D-lib magazine.

1997. (online), available from <<http://www.dlib.org/dlib/april97/cornell/04payette.html>>, (accessed 2004-09-19).
- [60] 高久 雅生, 江草 由佳, 宇陀 則彦, 石塚 英弘. Z39.50 による書誌データ検索システムの構築—Dublin Core を共通スキーマとして—. デジタル図書館. No.16, 1999, p.97-106. (オンライン), 入手先 <http://www.dl.ulis.ac.jp/DLjournal/No_16/12-masao/12-masao.html> (参照 2004-09-19).
- [61] Yahoo!. (online), available from <<http://www.yahoo.com/>>, (accessed 2000-06-21).
- [62] 須栗 歩人. “4-12-6 text widget のその他の機能”. 入門 Tcl/Tk. 東京, 秀和システム, 1998, p.168-175. (ISBN 4-87966-795-1).
- [63] UKOLN. Directory of Z39.50 targets in the UK. updated 1998-07-17, (online), available from <<http://www.ukoln.ac.uk/dlis/zdir/>>.
- [64] 阪口 哲男. Internet における情報資源の利用と提供の実例. デジタル図書館. No.1, 1994, p.43-50. (オンライン), 入手先 <http://www.dl.ulis.ac.jp/DLjournal/No_1/saka/saka.html>, (参照 2004-09-19).
- [65] “Report from January 2000 ZIG Meeting Working Group Session “Z39.50 and the Web””. Z39.50 Maintenance Agency. 2000-03-06, (online), available from <<http://lcweb.loc.gov/z3950/agency/zig/meetings/texas/zweb-report.html>>.
- [66] Hammer, Sebastian; Favaro, John. Z39.50 and the World Wide Web. D-lib magazine. 1996. (online), available from <<http://www.dlib.org/dlib/march96/briefings/03indexdata.html>>, (accessed 2004-09-19).
- [67] Z.X Home. (online), available from <<http://habanero.nhm.ukans.edu/z.x/Default.htm>>, (accessed 2000-07-21).
- [68] 経済産業省. 平成 12 年度データベース台帳総覧 台帳編 (本文). 東京, 財団法人データベース振興センター, 2001, 1035p.
- [69] 経済産業省. 平成 12 年度データベース台帳総覧 索引編. 東京, 財団法人データベース振興センター, 2001, 251p.
- [70] 国立情報学研究所. データベース・ディレクトリ. (オンライン), 入手先 <<http://www.nii.ac.jp/ir/dbmember/dbdr-j.html>>, (参照 2003-12-08).
- [71] 日外アソシエーツ. 世界 CD-ROM 総覧日本及び海外 2001. 東京, 紀伊国屋書店, 2001, 1905p. (ISBN 4-8169-1644-X).
- [72] GALE DIRECTORY OF DATABASES. Detroit, Gale Research, 2000, 1336p. (ISBN 0-7876-3289-9).
- [73] 国立国会図書館. 国立国会図書館データベース・ナビゲーション・サービス Dnavi, (オンライン), 入手先 <<http://dnavi.ndl.go.jp/>>, (参照 2004-09-04).
- [74] Zahir, Sajjad; Chang, Chew Lik. Online-Expert: An Expert System for Online Database Selec-

- tion. Journal of the American Society for Information Science, Vol.43, No.5, 1992, p.340-357.
- [75] Morris, Anne. Online company database selection: an evaluation of directories and CIDA (an expert system). Journal of Information Science. Vol.20, No.4, 1994. p.260-269.
- [76] Kahle, Brewster; Medlar, Art. An Information System for Corporate Users: Wide Area Information Servers, Online, Vol.15, No.5, 1991, p.56-60.
- [77] Meng, Weiyi; Yu, Clement; King-lup, Liu. Building Efficient and Effective Metasearch Engines, ACM Computing Surveys, Vol.34, No.1, 2002, p.48-89.
- [78] Manber, Udi; Bigot, Peter A. "The Search Broker". Proceedings of the USENIX Symposium in Internet Technologies and Systems. Monterey, California, 1997-12, USENIX Association, p.231-239.
- [79] French, James C.; Powell, Allison L. Metrics for Evaluating Database Selection Techniques. World Wide Web. Vol.3, No.3, 2000, p.153-163.
- [80] 岩澤 まり子, 吉田 亜津美, 大下 祥司, 中西 陽子. 図書館におけるレファレンス経験の知識化. 情報知識学会誌. Vol.12, No.2, 2002, p.37-48.
- [81] Hill, Linda L.; Janee, Greg; Dolin, Ron; Frew, James; Larsgaard, Mary. Collection Metadata Solutions for Digital Library Applications. Journal of the American Society for Information Science, Vol.50, No.13, 1999, p.1169-1181.
- [82] Powell, Andy; Michael Heaney; Lorcan Dempsey RSLP Collection Description, D-Lib Magazine, Vol.6, No.9, 2000. (online), available from <<http://www.dlib.org/dlib/september00/powell/09powell.html>>, (accessed 2004-09-19).
- [83] Bray, Tim; Paoli, Jean; Sperberg-McQueen, C. M. Extensible Markup Language (XML) 1.0. World Wide Web Consortium, 1998, REC-xml-19980210, (online), available from <<http://www.w3.org/TR/1998/REC-xml-19980210>>, (accessed 2003-12-08).
- [84] Apache Software Foundation. Apache HTTPD Project. (online), available from <<http://httpd.apache.org/>>, (accessed 2003-12-08).
- [85] Wall, Larry; Christiansen, Tom; Schwartz, Randal L. プログラミング Perl 改訂版. 近藤 嘉雪訳. 東京, オライリージャパン, 1997. 759p.
- [86] Sleepycat Software. The Sleepycat Software Homepage. (online), available from <<http://www.sleepycat.com/>>, (accessed 2003-12-08).
- [87] Veillard, Daniel. The XSLT C library for Gnome, (online), available from <<http://xmlsoft.org/XSLT/>>, (accessed 2003-12-08).
- [88] Open Source Technology Group. SourceForge.net. (online), available from <<http://sourceforge.net/>>, (accessed 2004-09-06).
- [89] Open Source Technology Group. freshmeat.net. (online), available from <<http://freshmeat.net/>>, (参照 2004-09-06).

- [90] istyle. @コスメ, (オンライン), 入手先 <<http://www.cosme.net/>>, (参照 2003-12-08).
- [91] カカクコム. 価格.com, (オンライン), 入手先 <<http://www.kakaku.com/>>, (参照 2003-12-08).
- [92] amazon.com. amazon.co.jp. (オンライン), 入手先 <<http://www.amazon.co.jp/>>, (参照 2003-12-08).
- [93] Leuf, Bo; Cunningham, Ward. The Wiki Way: Quick Collaboration on the Web. Addison-Wesley, 2001. 435p.
- [94] 関心空間運営事務局. “関心空間について”. 関心空間. (オンライン), 入手先 <<http://www.kanshin.com/index.php3?mode=about>>, (参照 2003-12-08).
- [95] OCLC Online Computer Library Center. (online), available from <<http://www.oclc.org/>>, (accessed 2003-12-08).
- [96] 国立情報学研究所開発・事業部コンテンツ課. NACSIS-CAT/ILL 目録所在情報サービスホームページ, (オンライン), 入手先 <<http://www.nii.ac.jp/CAT-ILL/contents/home.html>>, (参照 2003-12-08).
- [97] 武田 英明, 大向 一輝. Weblog の現在と展望 — セマンティック Web およびソーシャルネットワークの基盤として —. 情報処理. Vol.45, No.6, 2004, p.586-593.
- [98] Six Apart. “TrackBack Technical Specification”. Movable Type Publishing Platform. (online), available from <<http://www.movabletype.org/docs/mttrackback.html>>, (accessed 2004-09-23).

謝辞

本研究を進める際の全ての過程において、多くの方々からご指導、ご支援を賜りました。ここに記して、深く感謝の意を表したいと思います。

主任指導教官の石塚英弘先生（筑波大学教授）には、ご多忙の中、頻繁に個人指導の時間をとっていただき、本研究全般に渡りご指導頂きました。また生活面でも心配りして頂きました。深く感謝致します。杉本重雄先生（筑波大学教授）には、副指導教官をお引き受け頂き、私の研究における適切な指摘や関連研究についての情報提供などを頂きました。また私の国際会議の発表時には発表前に声を掛けてくださり、大変リラックスして発表できました。感謝致します。阪口哲男先生（筑波大学助教授）には、副指導教官をお引き受け頂くとともに、研究だけでなく様々なご助言を頂きました。特に、先生の豊富な経験に基づく幅広く詳細な知識からくる情報提供のおかげで、快適な研究環境を整備することができました。感謝致します。

山本毅雄先生（国立情報学研究所教授）には、図書館情報大学在籍中のおりには、ゼミに参加していただき、様々な観点からのご意見や、関連研究における歴史など多くのご助言を頂きました。また、山本先生が授業やゼミで行なわれた、辞書を使わず、英語をとにかく前から口を出して訳していき、わからない単語は英語のまま読んでいくという英語の学習方法と、先生の「英語は何歳になっても上達することができる」というお言葉のおかげで、私の英語に対する苦手意識が少なくなり、英語の論文を読むことがずいぶん楽になるようになりました。感謝致します。

田畑孝一先生（筑波大学教授）、石川徹也先生（筑波大学教授）には、本学位論文の予備審査および本審査を通じて、的確なご指摘と多くのご助言を頂きました。感謝致します。

石塚研究室のゼミに参加頂いた皆様や、情報メディアシステム分野所属の教官および学生の皆様には、私の研究に関する様々なご助言、指摘を頂きました。また、他の方の研究活動を間近で拝見し、議論に参加できる機会を得ることができましたことは私にとってかけがいのない財産となりました。感謝致します。

特に、図書館情報大学の学部入学以来の同期、高久雅生さんとは、共同研究の中で多くの議論を交わしました。これらの議論の中で様々なアイデアや問題点の解決方法が浮かんできたり、問題の整理ができました。また、多くの有用なソフトウェアを整備していただき、快適に研究を行うことができました。深く感謝致します。

宇陀則彦先生（筑波大学助教授）には、学部３年次からの研究を始めた当初に、右も左もわからず困っている私に、プレゼンテーションの仕方や論文の書きかたについて、丁寧にかつ粘り強く教えてくださいました。また、私が研究を続けていくきっかけを作ってくださいました。感謝致します。

長谷川秀彦先生（筑波大学助教授）には、ドキュメントを整備して共有することの大切さ、良い文章の書きかた、良いプレゼンテーションの仕方などを教えてくださいました。また、研究環境の整備や学生生活における注意事項など様々な助言を頂きました。感謝致します。

最後に、遠く広島からつくばに出ることを許してくれ、暖かく見守ってくれた両親に深く感謝致します。

発表文献一覧

査読制度のある学術雑誌に掲載の論文

- 江草由佳*, 高久雅生, 石塚英弘. 分散環境におけるデータベースに関する情報の共有システム. 情報知識学会誌, Vol.14, No.3, 2004, pp. 1–17.
- 高久雅生*, 江草由佳, 石塚英弘. Web サービスによる用語体系データの提供とその応用システム. 情報知識学会誌, Vol.14, No.1, 2004, pp. 11–22.
 - 高久雅生*, 江草由佳, 伊東千夏, 石塚英弘. データの再利用が可能な複数用語体系ブラウジングシステムの構築. 情報知識学会誌, Vol.13, No.1, 2003, pp. 10–22.
- 江草由佳*, 高久雅生, 宇陀則彦, 石塚英弘. Z39.50 データベース選択支援環境. 情報知識学会誌. Vol.11, No.2, 2001, pp. 1–10.
- 宇陀則彦*, 江草由佳, 高久雅生, 石塚英弘. Z39.50 による日本語書誌データ検索システム. 情報知識学会誌. Vol.9, No.2, 1999, pp. 1–15.

査読制度のある国際会議録に掲載の論文

Yuka EGUSA*, Masao TAKAKU, Hidehiro ISHIZUKA. Network Community Oriented Information Sharing System for Databases. Proceedings of International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society 2004 (DLKC'04), 2004-03, Tsukuba, pp. 64–71.

その他

- 伊東千夏*, 高久雅生, 江草由佳, 大懸晶子, 宇陀則彦, 石塚英弘. 用語体系の統合及び可視化システムの試作：用語体系の補完及び相違点の発見を目的として. 情報処理学会 情報学基礎研究会 研究報告. Vol.101, No.65, 2001, pp. 1–8.
- 江草由佳*, 高久雅生, 宇陀則彦, 石塚英弘. 情報共有による Z39.50 データベース選択支援

環境. 第 9 回研究報告会講演論文集. 情報知識学会, 東京, 2001-05. 情報知識学会, 2001, pp. 37-44.

- 高久雅生*, 江草由佳, 宇陀則彦, 石塚英弘. Z39.50 による書誌データ検索システムの構築 : Dublin Core を共通スキーマとして. 第 16 回 デジタル図書館ワークショップ. つくば, 1999-11. デジタル図書館. No.16, 1999, pp. 97-106.
- 江草由佳*, 真野泰久, 宇陀則彦, 石塚英弘. 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. 第 6 回研究報告会講演論文集. 情報知識学会, 東京, 1998-05. 情報知識学会, 1998, pp. 29-36.