

Z39.50 におけるデータベース選択支援機能の実現

The implementation of facility which supports database choice in Z39.50

学籍番号：98505

氏名：江草 由佳

本研究は Z39.50 検索クライアントを開発するとともに、検索要求にあったデータベースを選択するための支援機能を実現した。

Z39.50 は検索の際のクライアントとサーバ間で通信されるデータ構造やデータ処理方式について取り決めた ISO 及び JIS の規格である。Z39.50 に基づいてシステムを構築することにより、一つのクライアントから複数のサーバに接続できるため、利用者は異なるデータベースを透過的に検索できる。一方、World Wide Web (以下 WWW と呼ぶ) がインターネット上のハイパーテキストの概念に基づく分散型情報システムとして登場し、爆発的に普及した。現在、アメリカ議会図書館、OCLC をはじめ、アメリカやヨーロッパの図書館システムを中心に数多くの Z39.50 サーバが構築され、データベースは増加の一途をたどっている。そのため、Z39.50 によってデータベース選択の幅は広がったものの、検索要求にあったデータベースを見つけることは困難になった。

利用者が適切なデータベースを選択できるためには、データベース接続情報を網羅的に収集し、組織化してクライアントに登録する必要がある。しかしながら、データベース接続情報を収集し公開している WWW サイトや既存クライアントは、どれもデータベース接続情報を独自の形式で記述しているため、それぞれの形式を理解しなければならず、その大部分はデータベース名ごとに並び替えてあるだけなので、利用者自身で組織化しなければならない。また、全てのデータベース接続情報をひとつのクライアントが収集し組織化することは難しい。

この問題は利用者がそれぞれ収集し組織化したデータベース接続情報をお互いに補完し共有できるようになることで解決すると考えた。そこで、本研究では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 WWW に着目し、データベースに接続するための情報を WWW で共有し、Z39.50 でデータベースにアクセスし、検索するシステムを開発した。

本システムの構成要素は以下の 3 点である。

1. Z39.50 用 URL によるデータベース接続情報
2. Z39.50 用 URL を扱える WWW ブラウザ
3. その WWW ブラウザから呼び出せる Z39.50 クライアント

Z39.50 用 URL は、RFC で提案された Z39.50 サーバを一意に表すための URL である。Z39.50 用 URL でデータベース接続情報を記述することにより、全てのデータベース接続情報を同一形式で記述でき、かつデータベースを一意に識別できるため、WWW ブラウザからデータベースに直接アクセスすることが容易になる。しかし、Z39.50 用 URL を解釈できる WWW ブラウザは確認出来なかったため、本研究では新たに WWW ブラウザを開発した。この WWW ブラウザと Z39.50 クライアントにより、利用者は通常の WWW ページを閲覧したり、リンクをたどるのと同じ操作で、Z39.50 データベースを検索できる。具体的には、リンク集やディレクトリ型検索エンジンを利用して WWW ページを閲覧するのと同様に、すでに収集し組織化されたデータベース接続情報を利用して検索できる。

また、本研究ではデータベースに関する情報を収集し WWW 上で公開しているサイトから HTML 文書を取り寄せ、その記述を解析してデータベース接続情報を抜き出し Z39.50 用 URL に変換した。さらに、データベース接続情報の組織化も行った。組織化としては頻繁に利用するデータベースリストの作成や、分野ごとにデータベース接続情報を自動分類した。分野ごとの自動分類はデータベースのインデックス語とヒット件数を取り寄せる Z39.50 の Scan 機能を利用して行った。自動分類の利点は人手で分類するよりも、素早くデータベースを分類し追加できる点にある。

本研究では、Z39.50 と WWW の特性を活かした新しいシステム形態の可能性を示したとともに、個別に収集し組織化されていたデータベース接続情報を収集・加工・流通・利用というサイクルの中で共有できるようになった。

I implemented a Z39.50 client containing a facility to support database choice for users' demand for a search.

Z39.50 is an ISO and JIS standard, and specifies a client/server based protocol for information retrieval. It specifies procedures and structures for a client to search a database provided by a server, to retrieve database records identified by a search, and to scan a term list. It is possible to search transparently more than one Z39.50 server from one client. On the other hand, World Wide Web, which is based on the idea of the hypertext distributed information system on Internet, is spreading explosively. A lot of Z39.50 servers has been built including the Library of Congress, OCLC around the library system in United States and Europe, and databases fall into the increase. Because of that, though the width of the database choice spread out by Z39.50, it became difficult to find a database in a demand for a search.

Information for database connection is collected exhaustively and is organized, and users must register the information for the client in order to choose the suitable database. There are WWW sites which collect the information for database connections and introduced to the public. However, because all the information for database connections described with the individual form in these site and clients, users must understand each form. Moreover, users must organize the information for database connections by a user himself because the most part of the WWW site and the client is just arranged in database name of the order. Moreover, it is difficult for only one client to collect all information for database connection and to organize it.

These problems can be solved by complementing the information for database connection that a user collected them respectively and which was organized to each other and sharing it. In this research, it paid attention to WWW that each other could share the contents which an individual and organization made. This system which the information was shared with WWW, and database were accessed and searched with Z39.50 was developed.

The components of this system are three points of the following.

1. Information for database connection by the URL for Z39.50
2. The WWW browser that can support the URL for Z39.50
3. The Z39.50 client, which can be launched from that browser

The URL for Z39.50, proposed with RFC, is a mechanism to identify a unique Z39.50 server. Using the URL for Z39.50, all information for database connection can be described in the same form and can be uniquely identified a database. So it becomes easy to access a database directly from the WWW browser. But, it was developed by this research newly because a WWW browser that could support URL for Z39.50 didn't exist. With this WWW browser and the Z39.50 client, users can search Z39.50 databases by the operation which is the same as browsing usual WWW pages and following links. For example, as users browse WWW pages with a link page and a directory-type search engine, users can search databases with information for database connection to collect already and which is organized.

Moreover, it obtained a HTML document by this research from the WWW site where information about the database was collected. Then, that description was analyzed, and information for database connection was picked out, and it converted into the URL for Z39.50. Furthermore, the organization of the information for database connection was done, too. The gathering of a database list to use frequently and the automatic classification of the information for database connection information of every field were done. The automatic classification of every field used the Scan facility on Z39.50 that can fetch ordered index words of the database and the number of hit. The advantage of the automatic classification is quick creation of the list rather than it is classified by hands.

I showed that there is a strong potential for a profitable and synergistics relationship between the WWW and Z39.50. And by this research, it became to share information, which respectively collected and organized, in the cycle of the collection, processing, circulation, and use.

(指導教官：石塚 英弘)