

汎用的なsybperlを用いた図書館データベースとWWWの連携 — ユーザー指向の図書検索システムが備えるべき要件とは —

遠藤 教昭*、竹谷隆則**、進藤 浩一*

* 岩手大学・人文社会科学部・環境情報科学コース
**岩手大学附属図書館

本研究では、全文検索エンジン、RDBMS、およびWWWを連携させることによってユーザー指向を心掛けた図書検索システムを試験的に構築し、そのようなタイプの検索システムが備えるべき要件に関して考察した。その結果、WWWをフロントエンドに用いたユーザーフレンドリーな設計、瞬時的に応答すること、キーワード検索が可能であること、キーワード抽出用の辞書が充実していること、複数キーワードによる柔軟な検索が可能であること、キーワードの同義語、類義語による検索にも対応できることなどが、ユーザー指向の図書検索システムに必要な条件と結論できた。

Requirements for User Oriented Library Searching System

Noriaki ENDO*, Takanori TAKEYA** and Koichi SHINDO*

* Course of Environmental and Information Science, Iwate University HSS
** The Library of Iwate University

We analyzed the requirements for user oriented library searching system by evaluating the pilot system we constructed using WAIS(full text searching system), RDBMS and WWW. The conclusion was that user friendly design using WWW frontend, immediate response for keyword searching, suitable dictionary for keyword extracting, ability for multi keyword searching, searching ability using thesaurus, and so on were the requirements for the system of this type.

1. はじめに

1996年のはじめには、全国の大学図書館の中でWWW(World Wide Web)上での検索をサポートしているところはほとんどなかった。その頃の図書検索システムにおいては、まだ従来のように、telnet接続によってコマンドラインで検索するのが主流であった。

しかし現在では、かなり多くの大学でWWW版の図書検索システムが公開されている。そこで、著者もできるだけ使用してみたが、利用者が満足する状態とは必ずしも思えなかった。例えば、各フィールドにおいての前方一致検索しかサポートしていないとこ

ろが多いし、部分一致検索あるいはキーワード検索をサポートしていても速度が遅いなどの問題点があった。また、検索結果が意図したものとは異なるとか、検索フォームが見た目に複雑すぎる場所が多いという印象も受けた。

そのような状況を鑑み、本研究では、著者が全文検索エンジン、RDBMS(Relational Database Management System)、およびWWWを連携させることによってユーザー指向を心掛けた図書検索システムを試験的に構築し、その結果をもとにユーザー指向の図書検索システムが備えるべき要件に関して考察を深め

ることを目的とした。

2. ユーザー指向の図書検索システムを目指し考慮した事項

ここで述べるユーザー指向のシステムというのは、エンドユーザーがストレスなく目的を達せられるシステムのことを意味している。

従来のコマンドベースの図書検索システムでは、複数の検索結果を論理演算して最終の結果を得るのが普通である。これはけっこう煩雑な作業であり、コンピュータに慣れていない初心者はもちろん、ときどき検索を行う一般ユーザーにとってもあまり使いやすいシステムではない。

それに比較して、最近主流の WWW ベースの図書検索システムは、試行錯誤で検索条件を追加あるいは削除しながら検索できるので使い易く、ユーザー指向のシステムと言える。ただそのようなシステムにおいても、検索速度が遅いと、検索条件を変更しながら結果を絞りこむのが、やはり苦痛を伴う作業になってしまう。

著者は、真にユーザー指向のシステムを目指すには、検索ボタンを押したらほぼ瞬間的に検索が終了するのが望ましいのではないかと考えている。また、書誌名や著者名に含まれる単語で自由な検索ができなければならないと思う。そのため今回は、RDBMS 単独ではなく全文検索エンジンを併用して、的確なキーワード抽出と検索の高速化を図ることにした。また、初心者が抵抗を感じないように、検索フォームには必要十分な情報のみを表示するようにした。

3. システムの構築

1) 検索システムの概要

今回試作した図書検索システムでは、ユーザーが検索語を入れるとまず全文検索データ

ベースを検索して書誌番号と書誌名を表示し、その表示されたものの中から希望のものをクリックすると、RDBMS の書誌データベースを検索して、その書誌の詳細情報を表示するようにした(図1～図3)。

2) RDBMS と WWW の連携

RDBMS と WWW を CGI(Common Gateway Interface) で連携する方法、あるいは RDBMS や WWW サーバーソフトウェアのベンダーの独自 API(Application Programming Interface) で連携する方法がある。また、今後は JDBC(Java Database Connectivity)による連携が増えてくる可能性もある。

今回はフリーの WWW サーバーソフトウェアの NCSA httpd を用いているので、最も一般的で汎用性が高い CGI によって図書館データベースと WWW を連携させた。

3) RDBMS 用のゲートウェイプログラムの選択

本学の図書館データベースでは、RDBMS として Sybase 社の UNIX 版 Sybase SQL Server (以下 Sybase と記す) を使用している。

Sybase と WWW の連携に関しては、web.sql という Sybase 社の製品があるが、この製品は Sun Microsystems の Solaris2.3 を使用している本学の Sybase システムでは使用できなかった。

そこで、当初は NCSA (The National Center for Supercomputing Applications) のフリーソフトである gsql (C 言語で書かれた CGI プログラム) を候補に上げた。しかしながら、gsql は長い間バージョンアップがなく、最新の Sybase に対応させるには改造が必要なようであり、C 言語に精通していないと利用が難しい感じであった。また、改造するなら、

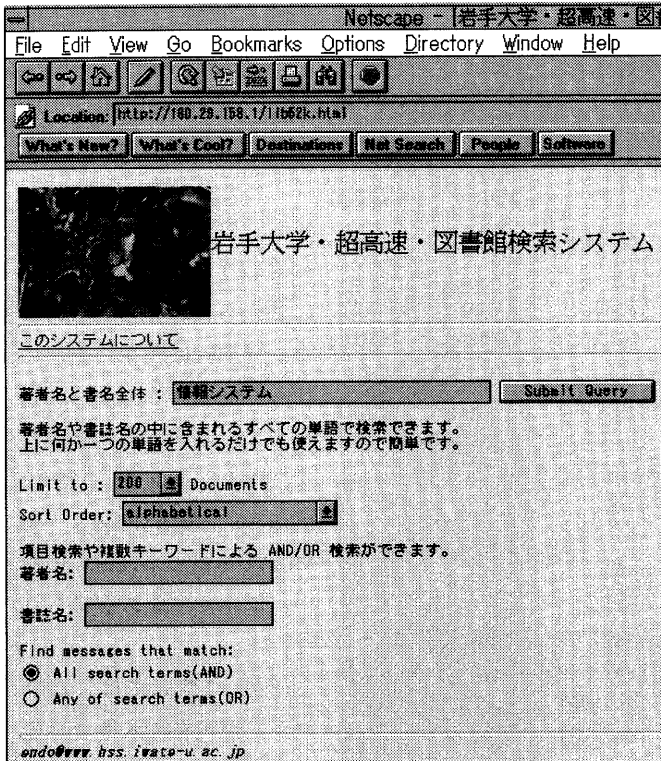


図1 検索開始画面

初めから自作するのと大差がないようにも思われた。

次に、Perl の参考書¹⁾で目にしたフリーソフトの `sybperl`(Perl 言語の Sybase 用拡張プログラム)²⁾を使用して見たところ、これが非常に使用しやすかった `sybperl` を用いると、通常 C 言語で用いる Sybase のアクセス用ライブラリを Perl 言語³⁾から使用できるようになる。Perl 言語自体も C 言語よりは気軽に使用できる言語であり、今回のような小規模なプログラムでは速度も十分に速い。親切な参考書⁴⁾も存在する。

以上のような経緯で、ゲートウェイプログラム作成には Perl + `sybperl` を採用することにした。

4) Sybase と WWW の連携法

まず Sybase が導入してある UNIX マシン

に `sybperl` をインストールする。次に、Perl 言語によって CGI プログラムを作成し、それを呼び出す HTML 文書と一緒に WWW サーバーに置く。CGI プログラムによって、Sybase との接続と SQL 文の実行が行われ、その結果が WWW ブラウザに HTML ファイルとして送られて来るようにする。

WWW クライアントから URL エンコードされて送られてくるデータをデコードするためには `cgi-lib.pl` を、漢字コード変換には `jcode.pl` を用いた⁵⁾。5) WAIS のシステム構築と WWW との連携

全文検索エンジンとしては、Z39.50 プロトコルに準拠する WAIS⁶⁾を用いた。WAIS にはフリーソフトウェア版⁷⁾と製品

版があるが、今回は製品版(富士通製、バージョンは最新版の V02L40)を用いた。これは ADJ 演算子を使用できるのが特長であり、複合語の検索に特に威力を発揮する。

具体的な WAIS データベースの作成法を次に述べる。まず、既存の Sybase の書誌データベースから書誌番号、編著者名、書誌名のフィールドを全てのレコードにわたって1つのファイルに書き出した。次に、そのファイルを用いて、項目検索が可能のように WAIS データベースを作成した。なお、現在の登録済み書誌数は約16万件である。

WAIS と WWW との連携に用いるゲートウェイソフトとしては、同製品に附属している WAISgateII を使用した。WAISgateII の細かい制御は、HTML ファイルに記載したパラメータによって行うようになっている。

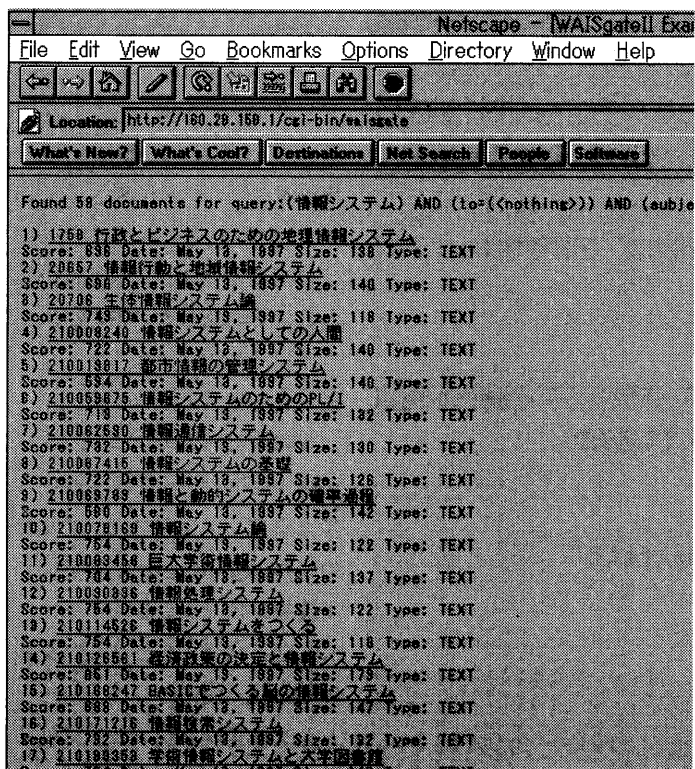


図2 検索結果の表示

6) システムの環境

WAIS用の環境

ソフトウェア WAIS V02L40, NCSA
httpd1.5

ハードウェア Sun Ultra Enterprise 2
(256MB Memory, UltraSPARC 200MHz×2,
Solaris2.5.1)

Sybase用の環境

ソフトウェア Sybase System10, NCSA
httpd1.5, Perl5.002, sybperl2.0

ハードウェア Sun SPARCstation 10
(128MB Memory, SuperSPARC 40MHz,
Solaris2.3)

4. 結果

システムの評価を中心に結果を述べる。

1) 検索速度

検索速度に関しては、パソコンで Netscape Navigator 3.0 と Internet Explorer3.02J を使って実験した。1クライアントで実験した場合は、WAIS の検索はほぼ瞬時的に終わり、Sybase の検索は1秒から数秒くらいであった。WAIS は現在、上記の Ultra Enterprise 2 によって運用しているが、当初試験した SPARCserver 20 (128MB Memory, SuperSPARC+ 60 MHz)でもほとんど遜色のないパフォーマンスが得られた。

Sybase に関しては、稼働マシンが現在の水準では能力が低めである上に、1台で多くの業務をこなしているため、繁忙時には結構時

間がかかることもあった。WAIS が速いだけに少し残念だが、1998 年には高速のサーバーにリプレースされるので、検索速度の向上に期待が持てる。

WAIS の検索に関しては、高速で快適そのものである。条件を変えて何回検索を繰り返してもストレスがたまることはないし、検索結果も意図したようなものが得られた。したがって、この部分に関しては、ユーザー指向の検索システムを作るといふ所期の目的を達せられたといえよう。比較的多くのユーザーが同時に利用しても問題はないと思われる。

WAIS の検索のボトルネックになる要因としては、ネットワークの混雑も考えられるが、大学内からのアクセスに関しては、ある程度の対策は考えている。最近各大学でも見られるように、岩手大学のネットワークは ATM

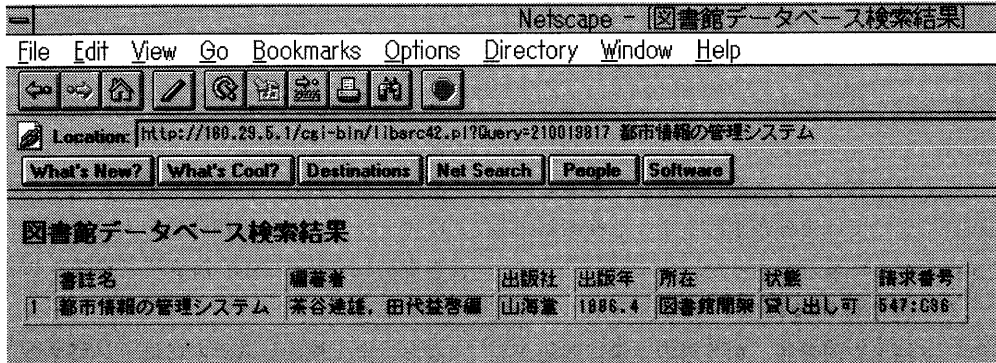


図3 検索結果の詳細表示

をバックボーンとしているが、Ultra Enterprise 2は100MbpsのEthernet(100Base-TX)でATM支線スイッチ直結のLAN Switchに接続している(Lan Emulation)。学生端末に関しても、本学部では一部は同様に100M Lan Switchに直結している。予算の関係で全部100M接続というわけにはいかなかったが、なるべくネットワークがボトルネックにならないように配慮している。今後はますますPC端末の100M化を進めるのが望ましい。

2) ブラウザに対する配慮

WWWサーバーにアクセスする場合、クライアント側の負担は通常さほど問題にならないが、大量の出力結果が出る場合はそうとも限らない。

ブラウザに対して、著しく大量のデータが一度に送られると、ブラウザがハングアップしてしまうことがある。従って、一度に表示される図書数の制限が必要であった。ただ、表示行数の制限は特にシステムの欠点とはならない。表示結果が多すぎるということは、あまり有効な検索とは言えないからである。

また、出力項目を画面に表示する場合に、見やすいように作表機能を使用した。これが結構クライアント側の負担になることがわかった。送られて来たデータを表に構成する

のがクライアント側のブラウザであることが原因である。前項に書いた表示行数の制限は、この負担の軽減にも役立っている。

5. 考察

1) 岩手大学附属図書館の正規のWWW図書検索システムについて

岩手大学附属図書館の正規のWWW図書検索システム(以下では既存システムと記す)は、UNIXベースでRDBMSとWWWを連携して構築されている。この検索システムは、一般的な水準に比較して見れば十分実用的なものであるが、改善すべき点も存在する。

このシステムでは、書名などの項目検索では、前方一致(および結果的には完全一致)しかサポートされていないが、そのかわり書誌のキーワードによる検索がサポートされている。したがって一般ユーザーは、キーワード検索を中心に利用していると考えられる。RDBMSで部分一致検索を行うと一般的に遅いので、この設計は評価できると思われる。

しかし、キーワードがどのように登録されているかが、システムの利用勝手を大きく左右する。既存システムでキーワードを登録する場合は、あるツールで著者名や書名から単語を切り出し、その単語の同義語や類義語も

あわせて該当の書籍に関するキーワードとして登録しているようである。つまり、書名に含まれない単語もキーワードとして機械的に定義されることになる。これは一見便利なようにも思われたが、実際に使用してみると、場合によっては自分が求めるような検索結果が得られない場合もあった。

その理由は明確にはわからないが、キーワード抽出や同義語、類義語の定義に問題があるのかも知れない。例えば、キーワード抽出に関して気付いた点を述べると、一文字の単語（例えば「犬」など）が抽出されていないことがあった。一文字の単語は結構多いので、この点に関しては今後のバージョンアップが必要であろう。

2) 今回構築したシステムで検索してみる

既存システムのキーワード検索に相当するのが、本システムでは、書名と著者名を合わせた全体から探すモードである。書名フィールド単独から探すよりも、この中から探すのが実際操作してみると意外とよかった。それは、著者名の中に、重要なキーワードが含まれることも多いからである。具体的には、何々編集委員会とか何々検討委員会とかの、何々のところに重要な語が含まれることがあった。

3) 今回構築したシステムと既存システムとの比較

今回試作したシステムではインデックス作成を別途行う必要があり、その作成中は検索ができないという欠点がある。インデックス作成に関しては、当日の追加データに関して、ユーザーのほとんどいない深夜に cron ジョブで自動的に作成するという方法もあるが、時間がかかるということには変わりはない。

ただ、今回のシステムの高速性や検索結果

の的確さは、何にも代えられない魅力といえよう。本システムの長所は、WAIS の長所、すなわち高速な検索、インデックス作成時の的確なキーワード抽出、検索の柔軟さなどによるところが大きい。検索の柔軟さとは、例えばキーワードとして複合語を入力した場合は自動的に各語が ADJ 演算子で接続されるなどである。

なお、全文検索システムにはインデックス作成の際にキーワード抽出を行わないも多いが、そのようなタイプのものではユーザーが意図しない結果が得られることもあるので注意を要する。たとえば、「アイヌ」で検索して「アイヌ」という単語が含まれる書誌が多数ヒットするなどということが考えられる。したがって、形態素解析機能を有する WAIS のシステムは評価に値すると思われる。

ただ、今後は WAIS でもユーザーが perl などからゲートウェイソフトを柔軟に制御できるようにバージョンアップされることを望みたい。現在は HTML ファイルでゲートウェイを制御するようになっていてそれなりに使い易いが、sybperl のようにもうすこし小回りがきくようになれば、更に利用価値が増すと思われる。

6. ユーザー指向の図書検索システムが備えるべき要件とは

これまで述べてきたことを総括して、著者が考えるユーザー指向の図書検索システムが備えるべき要件をまとめると、以下のとおりである。

1) WWW をフロントエンドに用いたユーザーフレンドリーな設計がされており、検索フォームはなるべくシンプルであること。

2) 高速な検索が可能で、検索要求には原則と

- して瞬間的に応答すること。
- 3)書誌名や著者名からの的確なキーワード抽出が行われていること。すなわちキーワード抽出の辞書が充実していること。
 - 4)複数キーワードによる柔軟な検索が可能であること。また、複合語で検索する場合に、検索もれが極力少なくなるように考慮されていること。
 - 5)ユーザーが入力したキーワードの同義語、類義語にも対応できること。

今回のシステムは、5)を除き上記の要件をほぼ満たしていると思われるが、既存システムにあり本システムにはない同義語や読み仮名による検索は、本システムにおいても今後なんらかの形で実現したいと考えている。

7. おわりに

本研究では、ユーザー指向の図書検索システムとはどのようなものかということに関して考察してきた。今後、よりユーザー指向を深めた図書検索システムが出現してくることを期待して結びとしたい。

なお、本システムは当研究室のホームページ⁸⁾において公開中である。

謝辞： Sybase の件で貴重なアドバイスをいただいた db_ml (データベースのメーリングリスト、代表者：三山勝久氏) の皆様に深く感謝の意を表します。

参考文献

- 1)河野真治： Perl 入門、アスキー、東京、153-160、1994
- 2)<http://www.mbay.net/~mpeppler>
- 3) <http://www.perl.com/perl/>
- 4)前田 薫、小山裕司、他： Perl の国からようこそ、サイエンス社、東京、1993

- 5)ローラ・リメイ著、武舎広幸他訳：続 HTML 入門、プレントイスホール出版、東京、1995
- 6)遠藤教昭、岡田 仁、進藤浩一他： WWW と全文検索システム WAIS によるシラバス検索システム、情報処理学会・利用者指向の情報システムシンポジウム論文集、107-114、1996
- 7)<http://www.kusastro.kyoto-u.ac.jp/~baba/wais/>
- 8)<http://www.hss.iwate-u.ac.jp>

連絡先：

〒 020 岩手県盛岡市上田 3-18-34
岩手大学・人文学部・情報科学
遠藤 教昭
e-mail:endo@msv.cc.iwate-u.ac.jp
URL <http://www.hss.iwate-u.ac.jp>

