学術情報センター紀要 第7号 (1995)

研究論文

インターネットに適応した全文データベース検索システムの構成

Configuration of a Full Text Database Retrieval System for Internet Environment

学術情報センター　大山　敬三

Keizo OYAMA

National Center for Science Information Systems

要旨

本論文では現在のインターネット上での情報サービス環境を最大限に利用しつつ、高度で使用しやすい全文データベース検索サービスを提供できるシステムの構成方法を述べている。WWWのクライアントであるMosaicとサーバであるHTTPDを用い、既存の高度な検索エンジンとの間を独自のゲートウェイにより結合する。階層構造を持つ文書データを対象に柔軟で効率的な検索を可能とするインタフェースを実現している。

ABSTRACT

This paper describes a configuration method of a system which can provide full text database retrieval services. The system is easy to use and has advanced features, while utilizing the information services environment currently available on the Internet at most. It adopts a client software Mosaic and a server software HTTPD for the WWW service as a basis, and combines an existing advanced text retrieval engine with them through an original gateway. It realizes a user interface through which users can access to document data with hierarchical structures flexibly and efficiently.

[キーワード]　全文データベース、情報検索、文書構造、インターネット、ゲートウェイ、グラフィカルユーザインタフェース

[Keywords]　full text database, information retrieval, document structure, Internet, gateway, graphical user interface, HTTP, Mosaic, Pat, SGML

インターネットに適応した全文データベース検索システムの構成

1. はじめに

インターネットの整備とワークステーションの普及に伴い、インターネットを通じて様々な情報サービスが提供されており、その数は急激に増えつつある。これらの多くは情報収集・提供のための組織的な支援体制を伴わないため、内容やサービスの質に問題を抱えている例も多く、情報を探索し利用する立場からは必ずしも望ましくない状況も発生している。このため、インターネットを通じた情報サービスに批判的な意見も散見される。

しかし、一方で、従来から独自のネットワークや方式で提供してきた情報サービスを、インターネット経由でも利用可能にし、さらにはインターネットをサービスの中心に据える例も少なくない。これらの多くは従来からの膨大な蓄積を、より効率的に、より利用しやすく、できるだけ多くのユーザに提供しようとするものである。このような努力は、従来からのユーザにとってのみでなく、潜在的なユーザにとっても歓迎すべきものと言えるであろう。

筆者が所属する学術情報センターでも、従来からいわゆる情報検索サービスを提供している[1]が、これまであまりインターネットの機能を有効に利用できないでいた。その原因の一つとしては、サービス提供のためのホストシステムがいわゆるメインフレームコンピュータであり、インターネットで標準的に利用されている環境が適用できなかったことがあげられる。また、センターのユーザの多くが自由にインターネットを使える状態になかったためにセンターが積極的にインターネット対応を推進する動機を欠いたことも一つの理由としてあげることができよう。

しかし、ここ数年間で大学や研究機関における状況は大きく変わり、希望するものは誰でもインターネットを利用できるようになった。このような状況の変化に対し、情報サービスの提供機関は学術情報センターに限らず、ユーザの利便性を第一に考え、できるだけ迅速に適応していくことがきわめて重要であると筆者は考えている。本論文では、このような目的を満たすために筆者が中心になって研究開発を進めているインターネット上の情報サービスシステムの構成方法について述べる。

2. 情報サービスのツールの現状と利用方針

筆者はインターネットを介したサービスを、既存のサービスの代替ではなく、今後のサービスの中心的役割を担うものと考えている。そのため、既存のサービスが備えている機能の見直しを含め、新しい環境に適応した高度な機能を実現することが重要である。

一方で、インターネットはさまざまなボランティアが提供してきた多くのパブリックドメインのソフトウェアと、それを元にして開発された商用の製品群によって大きく発展してきた。このような発展形態は今後も継続するものと考えられる。

本研究においては、インターネット上で一般的に利用されているツールを有効に組み合わせながら必要な機能を実現していくことを基本的な考え方とする。特にユーザ環境としては、広く利用可能な一般的なツールを用いてサービスを利用できることを最重要の課題と考えている。そこで、まず本章では、インターネット上の情報サービスのためのツールの現状を整理し、本システムでの利用方針を検討する。

インターネット上の情報サーバとしてはarchie、WWW (World Wide Web)、gopher、WAIS等が一般的である[2]。これらはいずれもパブリックドメインのソフトウェアとし提供されており、さらにWWWとWAISについては機能や性能を強化した商用版も提供されている。これらはそれぞれ目的を異にし、実現可能なサービスにもそれぞれ特徴がある。全文データベース検索機能の実現という観点

からこれらを整理すると以下のようになる。

(1) archieはFTPで提供されているファイル名に対する文字列検索のみを提供する[3]。内容に対するサーチ機能はなく、情報へのアクセスも別途FTPを用いて行う必要がある。

(2) WWWはネットワーク上に分散した情報資源をハイパーテキストの概念に基づいて利用可能にするもので、HTMLという形式で文書間に予め張られたリンクをたどることによってアクセスを行う[4]。他の情報資源に対するリンクの設定も可能であるが、WWWサーバ自体には内容に対する検索機能はない。

(3) gopherはネットワーク上に分散した情報資源をメニュー形式で利用可能にするもので、予め階層構造に分類整理されたメニューをたどることによってアクセスを行う[5]。情報資源の一つの形態として内容に対する簡単な文字列サーチ機能を持つが、情報検索機能としてはきわめて不十分である。

(4) WAISはサーバ中に格納されている文書に対し、文書中に現れる自由語を用いて頻度や類似度に応じた情報検索機能を提供する[6]。情報検索の知識がないユーザでも比較的容易に使用できるが、履歴や集合演算の機能がないため、複雑な検索を試行錯誤的に行うには不十分である。

また、これらの情報サービスを利用するためのユーザインタフェースとなるクライアントには、GUIを用いて統合的に対話が可能なMosaicがパブリックドメインソフトウェアとして普及しており[7]、商用版も各種のものが利用可能となってきている[8] (以下、これらを総称してMosaicと呼ぶ)。Mosaicは元々WWWのビューワとして開発されたが、現在ではWWWサーバ (HTTPプロトコルによる) ばかりでなく、gopherサーバやWAISサーバ (Z39.50プロトコルによる) へのアクセスが可能であり、また、HTTPプロトコルによりゲートウェイを介してarchieサーバなど、他の情報サーバへもアクセス可能である。gopherやWAISには専用のクライアントもあるが、普及の度合いや使いやすさを考慮すると、Mosaicをクライアントとして利用するのがユーザにとって最も有効であろう。

一方、全文データベースの検索には、単に自由語による内容のサーチを可能とするだけでなく、文書の階層構造に即した検索と取り出しが可能であることが必須である。既存の情報検索システムの中では、カナダOpenText社製のPat[9]が機能的に必要条件をもっとも良く満たしている。また、今後、より高度なシステムが利用可能になることも期待される。本研究の目的はMosaicを通じてそれらを適時に利用可能にするための方式を確立することにある。

Mosaicから利用可能な標準的なプロトコルとしてはZ39.50プロトコルのサブセットであるWAISプロトコルと、WWWで用いられているHTTPプロトコルがある。Z39.50は情報検索用の汎用の標準通信プロトコルであり、セッションの保持機能があって比較的高度な検索が可能であるが、WAISプロトコルはこのセッション保持機能を省略している[10]。また、Z39.50では検索対象のデータの階層構造を利用した検索・表示要求に対応できない。

HTTPはHTML文書などのハイパーテキストを転送するためのプロトコルであるが、HTMLのFORMの機能を用いるとユーザに提示する検索用のGUIを送ることができる[11]。HTMLではセマンティクスは何も規定しないので、FORMを設計するときにユーザに容易に理解可能なように考慮することが重要である。FORMに入力されたデータはゲートウェイを通じて情報サーバに届けられ、情報サーバからの出力をFORMにフィードバックすることも可能であるので、Mosaicを通じたユーザとの対話処理用に適している。

HTTP自体はステートレスでありセッションの保持機能がないが、HTTPのゲートウェイの高度な構

成方式によりこの機能が実現できれば、Patなどの強力な全文検索エンジンと組み合わせることにより、Mosaicを通じて試行錯誤を含む高度な対話処理を可能とする全文データベース検索システムを構成することができる。

3. 全文データベース検索システムの要件

本システムでは、検索の対象を、学術文献 (数ページ) から技術マニュアル (数千ページ) 程度の、階層構造を持つ文書の全文データとする。必ずしも必要な条件ではないが、SGML (Standard Generalized Markup Language) で記述されているものと仮定する。データベースにはこのような相互に比較的独立な全文データが多数格納されている。

本章では、このような環境において全文データベース検索システムが扱うべき文書データの構造を分析し、多数の文書の中から所望の文書を探し出すためにユーザが必要とする機能を検討することにより、システムに要求される要件を明らかにする。

3.1 文書データの構造と検索要求

本システムが対象とする文書データは一般に図 1 のような階層構造を持つ。この他にも参考文献とその参照のようなリンク構造もあるが、多数の文書データに対して一括して検索する場合にこれが有効な手がかりとなることはあまり無いと考えられるため、検索対象は階層構造のみとする。

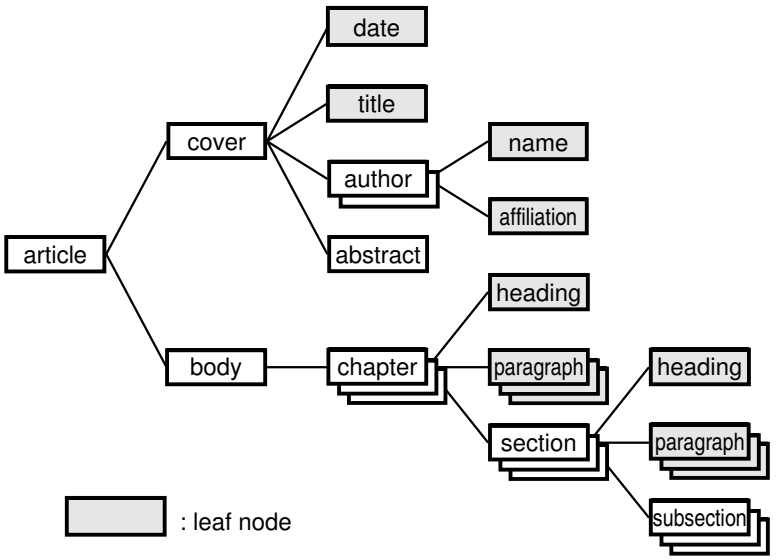


図1 文書の階層構造

SGMLで階層構造定義に関連して定義可能なものには、出現順などを指定する結合子と、繰り返しなどを指定する出現指示子がある。これらの中で、出現順などは多数の文書の中からキーワードを頼りに検索する環境ではあまり有効な利用方法は考えられない。一方、繰り返しは適切に取り扱えないと精度の高い検索が期待できない。

ユーザが必要とするであろうと考えられる検索条件の例としては、以下のようなものが考えられる。

(1) 節タイトルに「実験方法」が含まれ、節の本文中に「真空蒸着」が含まれるような節を持つ論文を表示する。
(2) 同一の段落に「バブル経済」と「経済摩擦」が含まれるような論文の書誌情報を表示する。
(3) 同一の段落に「バブル経済」と「経済摩擦」が含まれるような論文について、書誌情報と章・節のタイトル、およびその段落を、出現順に表示する。

これらのうち、(1)、(2)は検索対象と表示対象の関係が比較的単純であるため条件の記述も容易であるが、(3)はやや複雑となるためMosaicを通じてユーザが条件の指定をできるような簡潔なHTMLのFORMを作成することは困難である。本システムでは、検索対象と表示対象の関係が単純な包含関係にある場合のみを想定する。

3.2 システムの要件

情報検索システムの一般的な機能には大きく分けてサーチ機能と表示機能がある。全文データベース検索では通常の書誌型、索引型、あるいは抄録型の情報検索と比較して、以下のような機能が特に重要である。

(1) サーチ機能

(a) 文字列サーチ
ユーザが利用目的に応じてさまざまな角度から検索できるよう、テキスト中に出現する任意文字列でのサーチ、範囲検索、近接演算 (順序指定、順序非指定) などが可能でなければならない。特に日本語の学術情報では、辞書にない単語や複合語でのサーチができることも重要である。

(b) 論理演算
検索対象が階層構造を持っているため、階層に即した論理演算が必要であり、各階層における任意の文書要素および一致文字列に対して、論理積、論理和、論理差、否定などの論理演算が可能でなければならない。

(c) 集合演算
ユーザにとって、複雑な検索を最初から単一の論理式で記述することは困難であり、試行錯誤を繰り返す上でも過去の検索結果集合を組み合わせた集合演算が可能であることが重要である。論理演算同様、文書要素集合および一致文字列集合に対する共通集合、合併集合、排他集合、補集合などの集合演算が可能でなければならない。

(d) 階層演算
階層構造中の任意の文書要素を基点にしてサーチを行い、さらに異なる階層の文書要素に基点を移してサーチを行うためには、階層構造にある文書要素・文書要素間および文書要素・一致文字列間の包含と包摂の演算が、論理演算および集合演算の双方において可能でなければならない。

(2) 表示機能

(a) 出現位置の周辺表示
文中にある任意の文字列に対するキーワード検索を行う場合、その語の使われ方をユーザが予めすべて予想することは困難な場合が多いため、キーワードに一致した文字列がどのような文脈で現れたかを簡略に表示する、いわゆるKWIC表示が有効である。

(b) 文書要素を指定した表示
キーワードによるサーチやその結果の組み合わせにより得られた文書が適切であるかどうか判断し、また、それらの中から有用と思われるものを選択するためには、そのために必要かつ十分な部分を一覧できるような表示機能が必要である。たとえば文献情報部分のみを一覧表示させたり、文書ごとの目次 (文書タイトルと章節の見出しなど) を表示させたり、キーワードが見つかった段落を表示させたりできることが要求される。また、マニュアルなどでは該当する節を表示させるだけでユーザの目的を達せられることもある。

(c) 全文を簡易整形して表示する機能
選択した文書を通覧するためには、ある程度読みやすい形でオンライン表示できる必要がある。最終的にはSGMLで記述された原データを加工して厳密な整形出力をオフライン印刷できればよいであろうが、ユーザの利用性を高めるためには対話処理のなかで簡易整形表示できる機能が重要である。

これらの機能のうち、サーチ機能はサーバで実現する必要がある。表示機能はサーバ、クライアントのいずれでも実現可能であるが、本システムでクライアントに想定しているMosaicにはこれらを実現する機能がないため、サーバで実現する必要がある。また、仮にクライアントがこれらを実現する機能をもっていたとしても、通信の負荷を考慮すると多数ユーザの環境では現実的ではない。

全文データベース検索では特に文書がもつ階層構造を適切に処理する必要があるため、上記のような複雑な機能を持つシステムが必要であるが、一般のユーザがこのような機能を理解し、対象とする文書の具体的な構造を覚えて検索に有効に利用するのは極めて難しい。このため、メニューなどを用いることにより特別の知識を持たないユーザでも上記のような高度な機能を容易に利用できるようなユーザインタフェースを実現することが重要である。

4. システム構成

本章では、上記のような諸々の条件を満たすために採用した全体構成とその動作の概要を述べ、本システムの開発の中心となるフロントエンドについて詳述する。

4.1 全体構成

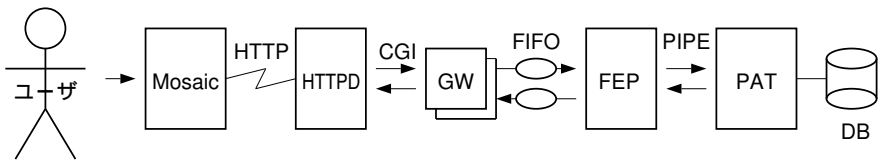
本システムでは前述の通り、クライアントとしてMosaicを利用し、WWWサーバを基礎として全文データベース検索サーバを構成する。全体構成を図(a)にす。各部分の機能と相互間の関係を簡単に述べる。

(1) HTTPクライアント

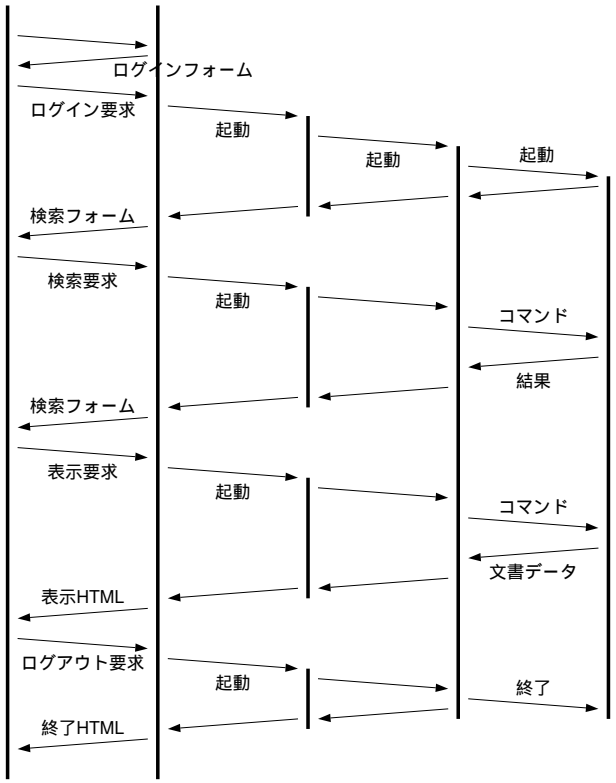
HTTPクライアントはユーザがシステムと対話するための直接のインタフェースを提供する。ユーザがインターネット上の情報サーバにアクセスするときに随時起動する。現在利用可能なものにはMosaic以外にもいくつかの種類が存在し、フォームを扱う機能を持つものであれば本システムのクライアントとして利用可能である。しかし、ユーザインタフェースの質や利用可能なプラットホームの多様性から、現在ではMosaicが最も利用しやすい。

(2) WWWサーバ

WWWサーバはクライアントからのアクセスに対して要求された情報を提供する。通常は予め用意されたHTML文書や画像ファイルを送信するだけであるが、本システムではフロントエンドで動的に



(a) 全体構成



(b) 動作の概要

図2 システムの概要

作成されるHTML文書をクライアントに送信する。全文データベース検索サーバの起動時に 1 回だけ起動され、システム停止まで存続する。クライアントとはHTTPプロトコルによりTCP/IPを介して通信する。ゲートウェイ側のインタフェースはCGI(Common Gateway Interface)として規格が定められている[11]。WWWサーバにはパブリックドメインや市販製品として利用可能な既製のものがいくつか存在する[8]が、今回はCERN版のHTTPDを用いることとした。

(3) ゲートウェイ

ゲートウェイはWWWサーバとフロントエンドの間のデータの仲介を行う。フォームが送られてくるごとにWWWサーバから起動され、標準入出力を通してWWWサーバと情報の授受を行う。フロントエンドとの通信は各フロントエンドと 1 対 1 に対応したFIFOファイルを通じて行う。結果の情報をWWWサーバに返すと消滅する。頻繁に生成と消滅を繰り返すのでシステムの性能を律する可能性

があるため、可能な限り機能を簡素化してある。

(4) フロントエンド

フロントエンドはユーザの要求を全文検索エンジンのコマンドに解釈して実行させるとともに、結果やデータの表示に必要なHTMLのフォームや文書を作成する。セッションの維持や検索の履歴の管理に必要な処理も行う。個々のセッションに対応して１つずつ動的に生成され、セッションが終了すると消滅する。全文検索エンジンとはパイプを通じて通信する。既存の構成要素で提供されていない機能の大部分はここで実現され、本システムのプログラム開発の大部分が含まれている。

(5) 全文検索エンジン

全文検索エンジンには、各種のパブリックドメインソフトウェアや市販製品を比較検討した結果、カナダOpenText社製のPatサーバを採用することとした。ただしこれは全文データベースに対する検索機能を評価した結果の選択であって、システムの構成上からの必要性によるものではない。従って、今後、より適当なものが利用可能になればそれを採用することも可能である。全文検索エンジンはフロントエンドにより管理されており、セッションが開設されるたびに起動され、セッションが終了すると消滅する。

4.2 動作の概要

全体の動作の概要を図2(b)に示す。ユーザはまずHTTPクライアントであるMosaicを起動し、ログインフォームをWWWサーバであるHTTPDから取り出す。フォームにユーザIDなどの必要な情報を記入し、使用するデータベースを指定してHTTPDに送る (ログイン要求)。HTTPDはゲートウェイプロセス (GWP) を起動してフォームに記入されたデータを渡す。GWPはセッションの環境を作成してからフロントエンドプロセス (FEP) を生成する。FEPは全文検索エンジン (Pat) を起動し初期化してから、検索フォームをGWP、HTTPDを経由して転送してMosaicに表示させる。

ユーザは、検索フォームに検索条件を記入してHTTPDに送り返す (検索要求)。HTTPDは新たなGWPを起動してフォームに記入された検索条件を渡す。GWPは対応するFEPを見つけて検索条件を渡す。FEPはこれをPatのコマンドに解釈してPatに渡す。Patから結果を受け取るとFEPは検索フォームに検索集合の一覧と内容の簡略表示を埋め込んで結果フォームを作成し、GWP、HTTPDを経由してMosaicに送り返し、表示させる。

結果フォームに簡略表示されたもののなかで詳細な表示をさせたいものがある場合は、ユーザは文書を選択してHTTPDに送り返す (表示要求)。HTTPDはGWPを起動して要求をFEPに渡し、FEPはそれをコマンドに解釈してPatに処理を行わせる。取り出した文書データはFEPがHTML形式に簡易整形してGWP、HTTPDを通じてMosaicに転送する。

一連の検索を終えると、ユーザは結果フォーム中に終了の指示を記入してHTTPDに送り返す (ログアウト要求)。GWPを通じてログアウト要求を受け取ったFEPは、Patを終了させ利用統計などを記した終了通知のHTML文書を作成し、これをGWP、HTTPDを経由してMosaicに転送した後に終了する。

4.3 フロントエンドの構成

フロントエンドは本システムの開発の中心であり、機能的にも最も複雑である。機能には大きく分けて、セッション管理機能、コマンド解釈機能、フォーム表示機能、文書表示機能がある。以下でこれらについて説明する。

(1) セッション管理機能

セッションの開始時には、ログインフォームに記入された情報からセッションIDを生成し、指定されたデータベースに対してPatを起動するとともに、以後のセッションの中でGWPからのアクセスに用いるFIFOポートを作成してからフォーム表示機能を呼び出す。セッションIDはユーザ名、クライアントホスト名、サーバホスト名、FEPのプロセスID、セッション開始時刻などから容易に想像できないように生成する。課金などのためにユーザ管理が必要な場合は、上記の処理を開始する前に、ログインフォームに記入されたユーザIDやパスワード、およびHTTPDから渡されるクライアントのホストIDなどを用いてユーザ認証を行う。

セッションの終了処理はユーザからのセッション終了要求を受けて行う。Patを終了させ、利用統計などを記入した終了通知を作成して返す。ユーザ管理が必要な場合は各種統計情報をファイルに書き出す。使用済みのFIFOポートを削除してプロセスを終了する。

HTTPにはセッションの概念がないため、クライアントとサーバは1回のアクセスごとに接続と切断を繰り返す。従って、クライアントがログアウト要求を送らないまま終了してしまうと、サーバはこれを検知できない。このため、FEPはセッションの保留時間を監視し、ユーザから要求がない状態が一定時間継続した場合は中断処理を行う。現在は中断処理では強制的に終了処理を実行しているが、サービスの内容に応じは、後で回復可能なように中断時のFEPとPatの状態を保存してセッションを解放するというような機能の導入も考えられる。

(2) コマンド解釈機能

検索フォームに埋め込まれた検索要求を取り出し、Patのコマンドに解釈して処理を実行させる。検索要求にはサーチ要求と表示要求がある。サーチ要求に対しては検索条件をPatのサーチコマンドに翻訳して実行させる。要求内容とPatが作成した検索集合を履歴として保持し、実行結果をフォーム表示機能に渡す。表示要求に対しては、選択された文書の中から指定された文書要素を取り出すためのPatのコマンドに翻訳して実行させる。取り出した文書データは文書表示機能に渡す。

(3) フォーム表示機能

あらかじめ対象データベースごとに用意された検索フォームのテンプレートに、サーチの履歴と検索集合の内容の簡略表示を埋め込んでHTTPDに渡す。履歴には最近の指定回数分の検索集合の番号、その件数、およびサーチ要求の内容を表示する。簡略表示には最新の検索集合に含まれる文書データの先頭部分を指定件数分だけ表示する。検索フォームのテンプレートにはサーチ文字列の入力フィールド、サーチ範囲を限定するための文書要素のメニュー、集合演算のための演算子のメニュー、階層演算のための文書要素のメニュー、表示要求で文書要素を指定するための文書要素のメニューなどが予め準備されている。履歴にはチェックボックスを付加し、集合演算のオペランドとして指定できるようにする。簡略表示にもチェックボックスを付加し、表示要求の文書指定をできるようにする。

(4) 文書表示機能

表示モードとして整形出力とソース出力を持つ。整形出力モードでは与えられたSGML文書をHTML形式に変換してクライアントに送る。変換のためのフィルタはデータベースごとに独立したプログラムとして用意しておく。ソース出力モードでは元のSGML文書をそのままプレインテキストとしてクライアントに送る。

5. ユーザインタフェース

階層構造を持つ文書を対象とする全文データベース検索システムのユーザインタフェースでは、文書の階層構造の処理を含む検索要求を入力できる必要があり、指定すべきパラメータの種類が多い。ユーザがこれらをすべて記憶し、適切に誤りなく直接入力することはきわめて困難である。そこで、本システムではこれらのパラメータをユーザに提示し、ユーザはそれらの中から選択してゆくだけで検索要求を入力できるようなユーザインタフェースを実現する。HTMLのフォームでMosaicに表示できるグラフィカルユーザインタフェースの構成要素には限界があり、ユーザの操作に対応したきめ細かい動作はできないが、基本的な検索の対話をするための要求は満たせると考えている。

図3、図4(a)、図4(b)および図5にユーザインタフェースの例を示す。図3はログインフォームの表示である。ユーザはユーザID、パスワードを入力し、データベースの一覧の中から 1 つを選択する。図4(a)、図4(b)は検索フォームの表示例である。サーチ要求の記入部分、履歴表示部分、簡略表示部分から構成されている。ユーザが直接入力するのはサーチ文字列だけである。図5は簡易整形した文書表示の例である。

図4(a)、図4(b)に示した検索フォームの操作について次に説明する。

(1) キーワードによる文字列サーチ

検索に用いる任意の文字列をフォームの「検索文字列」フィールドに入力する。検索範囲を限定する場合は「サーチ範囲」から対象とする文書要素を選択する。履歴表示部分に結果の集合番号、件数、および検索内容が表示され、簡略表示部分に該当部分のリストが表示される。表示する履歴の数と簡略表示の数はパラメータ設定部分の「履歴表示件数」と「簡略表示件数」で指定する。なお、この検索フォームでは論理演算はサポートしておらず、単一文字列によるサーチで集合を作っておいてから集合演算を行うことで代替する。

(2) 集合演算と階層演算

演算の種類を「演算」から選択し、オペランドを検索集合か文書要素の中から指定して「実行」する。演算の種類には集合演算、階層演算、近接演算がある。集合演算は 2 つのオペランドの文書要素の種類が一致しなければあまり意味がない。階層演算は 2 つのオペランドの文書要素の種類の間の上下関係が適切でないと意味の無い結果となる。近接演算は 2 つのオペランドが文字列サーチの結果でないと出現位置の間の距離を反映しなくなる。なお、距離 (文字数) はパラメータ設定部分の「近接演算距離」で指定する。

(3) 簡略表示

文字列サーチや集合・階層演算を行うと最新の検索集合の簡略表示が行われるが、以前の検索集合を見たいときには履歴部分の表示項目をクリックする。また、表示している部分の前後を見たい場合は上下の矢印のリンクをクリックする。

(4) 詳細表示

詳細表示には 2 つの方法がある。表示したい文書の簡略表示を直接クリックすると、その文書が詳細表示される。また、簡略表示の頭にあるチェックボックスをチェックしてから「表示」を行うと複数の文書を一括して詳細表示することができる。「全部表示」をチェックすると全ての文書を一括表示できる。表示の形式はパラメータ設定部分の「詳細表示形式」で選択する。

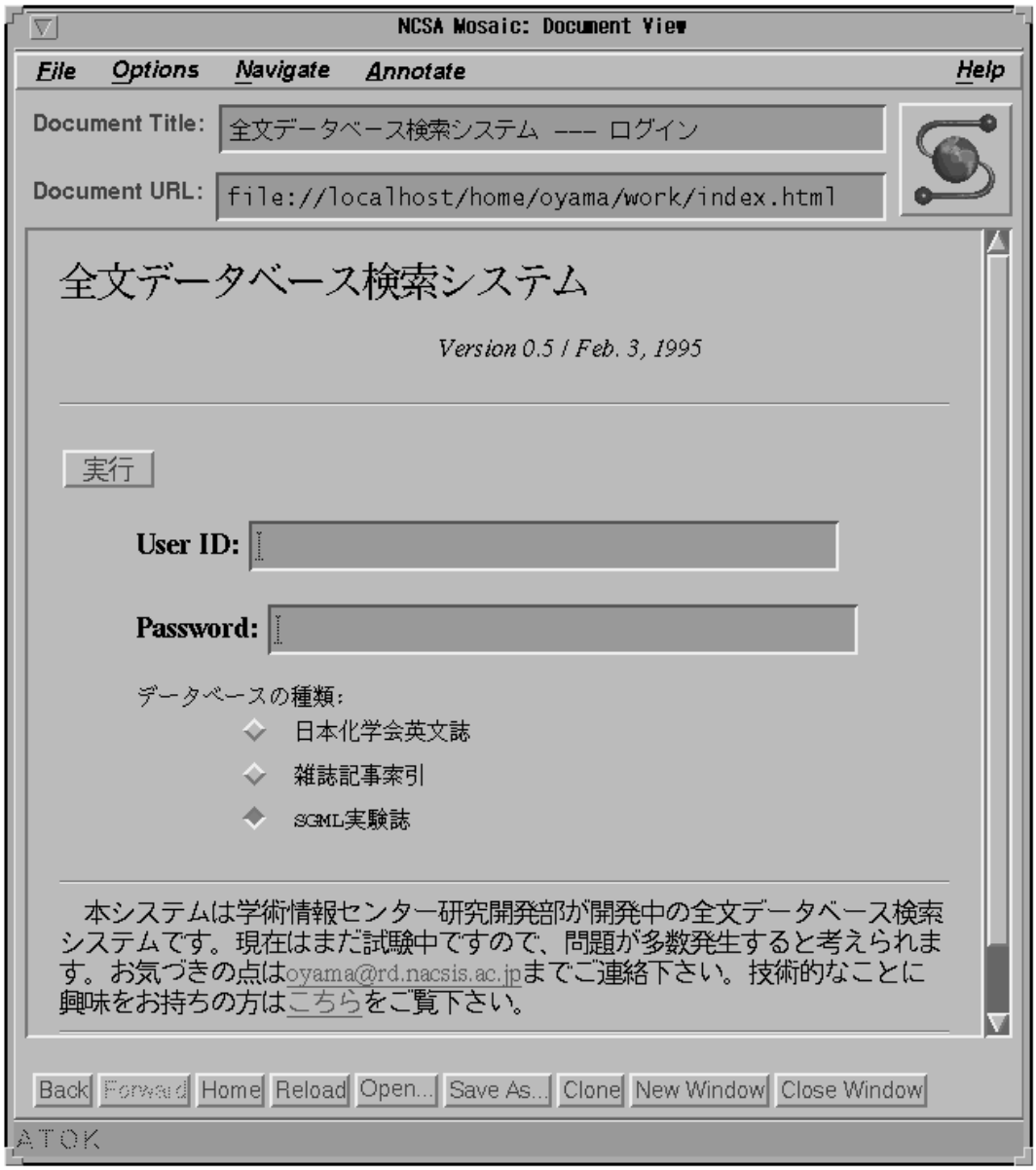


図3 ログインフォーム

(5) パラメータの設定

各種のパラメータの設定はパラメータ設定部分の項目を指定して「設定」を行う。

(6) その他

セッションの終了は「ログアウト」で行う。また、データベースの内容に関する情報と検索フォームの使用方法に関する情報はマニュアルのリンクをたどることで参照可能である。

図4-(a) 検索フォーム (その1)

6. 終わりに

本論文ではインターネット上でもっとも広く用いられているユーザ環境から利用可能な全文データベース検索システムの構成方法について述べた。このシステムは、データベースに収録されている文書の構造についての特別な知識なしにGUIを通じて容易に検索できることを特長とし、本格的で先進的な情報検索サービスの利用を可能とする。

本論文で述べた検索フォームは基本的なものであり、データベースの特性に応じて予めアクセス形態のモデル化が可能なものについては、それに沿った検索をより効率的に行うための検索フォームの

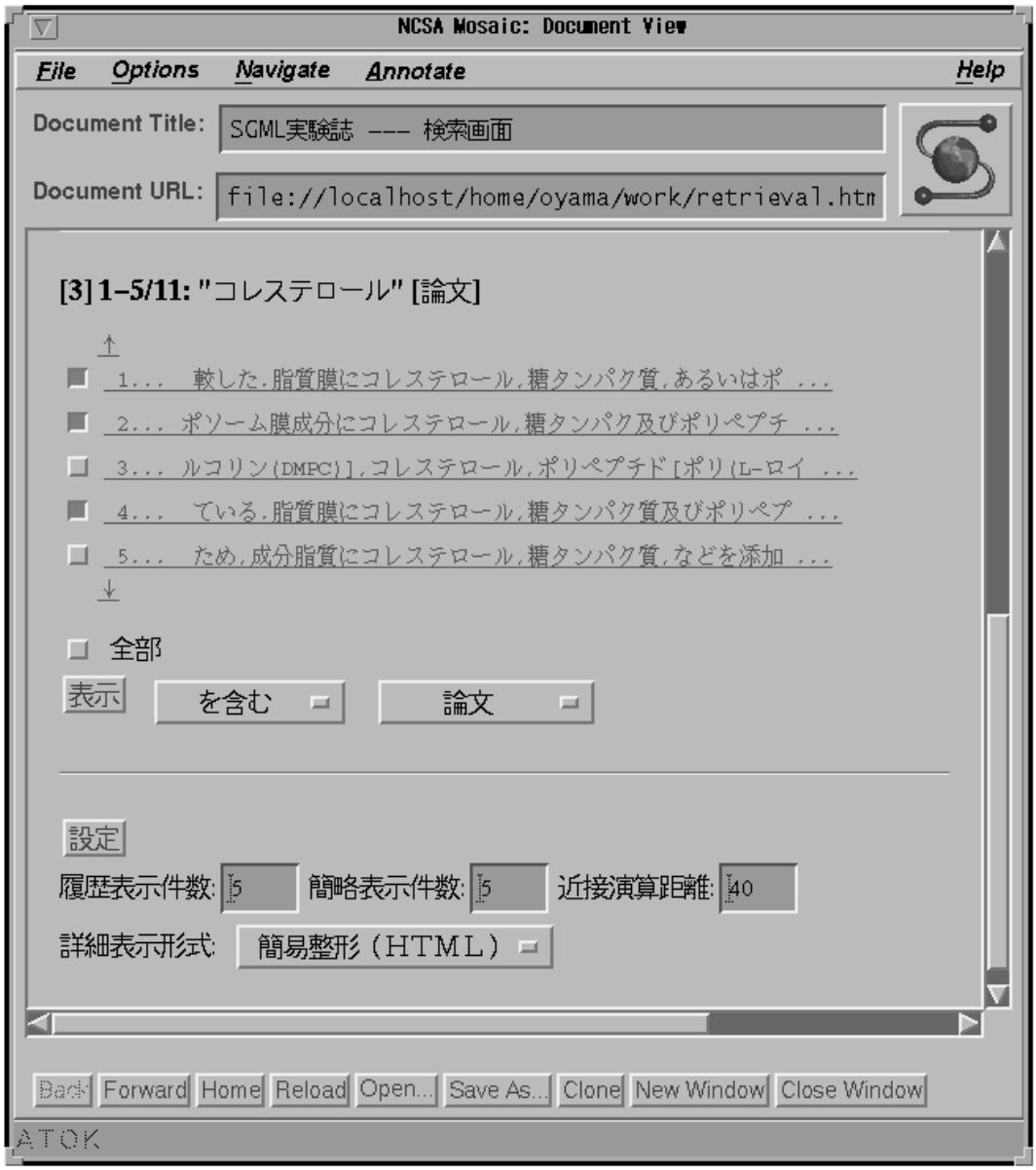


図4-(b) 検索フォーム (その2)

作成を行うのがよいであろう。複数の検索フォームの中から、適切なものを選んで使用することも容易に実現できる。

本構成方法は、より一般的な既存の情報検索サービスにも適用可能であり、利用方法が難しいためにユーザ層が限定されていたり有効な機能が十分に活用されていなかったような分野のサービスにも、より広く有効に利用される可能性を開くものである。

本システムは現在開発中であり、本論文で述べた内容は若干変更される可能性がある。完成後は利用実験による評価、改良を行った後にサービスを一般に公開する予定である。また、同様のサービス

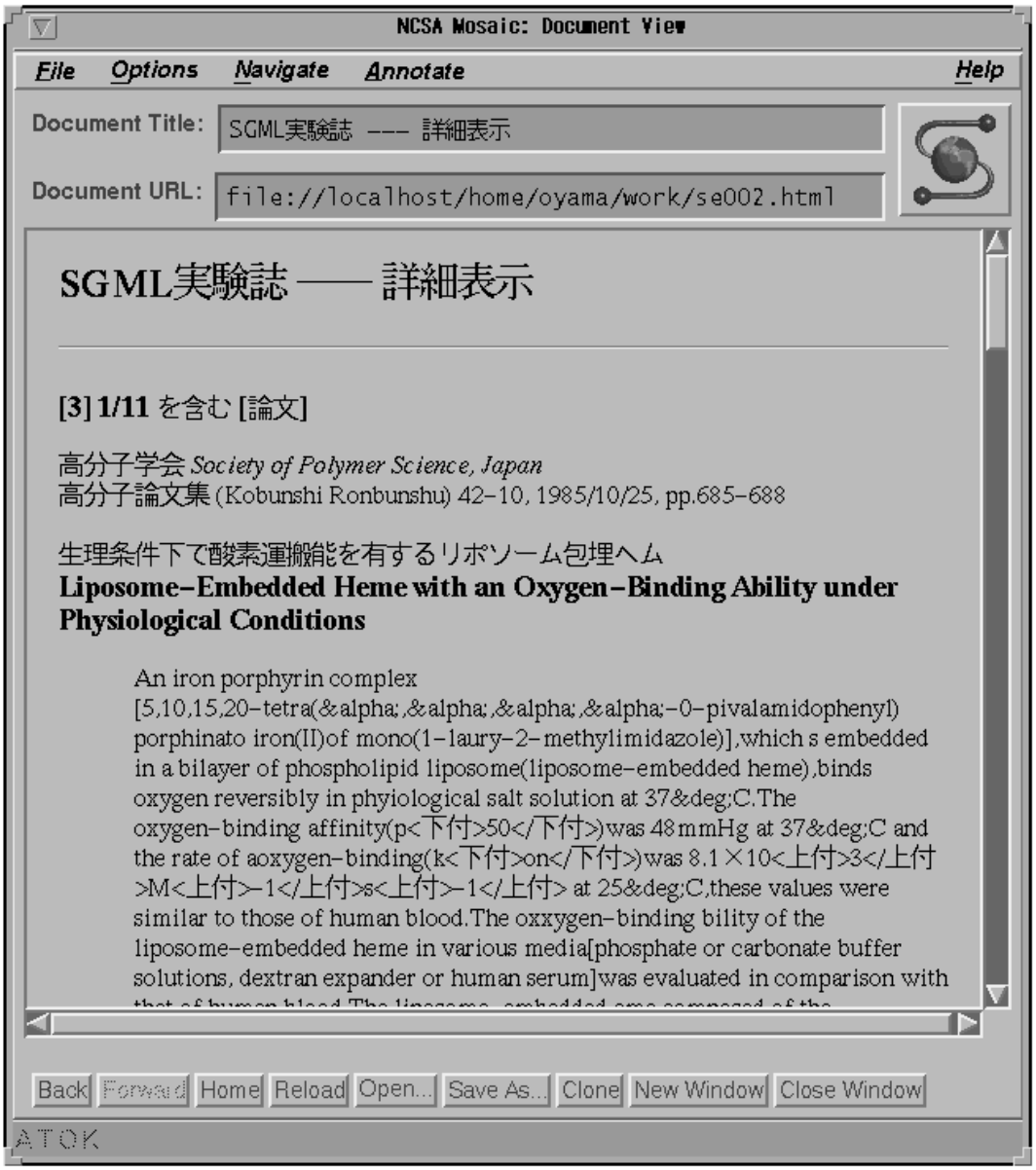


図5 詳細表示

を行おうとする者のために、開発したソースコード、HTMLフォームなども公開する予定である。

ただし、HTMLの記述能力やMosaicの表示能力の制約から、不満足な点が残っていることは否めない。履歴や簡略表示がスクロールできないのもそのひとつである。これらはHTMLの拡張やMosaicの機能強化により徐々に解決するであろうが、より専門的なユーザのためには専用のクライアントと通信プロトコルを用意すべきであろう。

学術情報センターでは現在、本システムの他に、文字型ユーザインタフェースの全文データベース検索システムを開発中であり、また、電子図書館の一部として機能する全文データベース検索サーバ

も開発中である。これらについては別の機会に紹介する。

参考文献

[1] 「情報検索サービス NACSIS-IR」, http://www.nacsis.ac.jp/ir/ir-j.html.

[2] 斎藤正史; 山口英,「インターネットの情報サービス」, 情報処理, Vol.34, No.12, pp.1415-1421, 1993.

[3] “Archie Reference Manual Pages”.

[4] “A Beginner’s Guide to HTML”, http://www.ncsa.uiuc.edu/demoweb/html-primer.html.

[5] “Gopher Reference Manual Pages”.

[6] “WAIS 2.0 Technical Documentation”, http://www.wais.com/newhomepages/techtalk.html.

[7] “NCSA Mosaic Home Page”, http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/NCSAMosaicHome.html.

[8] 「社内情報共有に新時代 — インターネットのWWWが急浮上」, 日経コミュニケーション, 1995.1.6, pp.76-97.

[9] “Pat Reference Manual”, OpenText Corp.

[10] “WAIS(tm) Network Publishing using Z39.50”, http://www.wais.com/newhomepages/z3950.html.

[11] “The Common Gateway Interface”, http://hoohoo.ncsa.uiuc.edu/cgi/overview.html.

[12] Goldfarb, Charles F., “The SGML Handbook”, Clarendon Press, 663pp. 1990.