

WDCMにおけるWWWサーバの構築

○宮崎 智*、志村 純子*
菅原 秀明*、一柳 芳浩**

The implementation of WWW server on WDCM

○Satoru MIYAZAKI*
Junko SHIMURA*
Hideaki SUGAWARA*
Yoshihiro ICHIYANAGI**

In this paper, we report some advanced technique for the implementation of WWW server. On WDCM Web server, we introduced WAIS program for searchable engine as background jobs and LINK to connect records beyond the databases by HTML language. This technique relates all our independent databases and allows users to traverse a few databases quick and smooth. We also developed an interface between Web server through MOSAIC and other useful programs, for instance, RDB, OODB and E-mail systems so on. "CGI POST METHOD" of HTML plays the leading part to implement the interface program.

はじめに

コンピュータネットワークとマルチメディア環境の担い手として、期待されているインターネット。現在その中心的な役割を演じようとしているのがWorld Wide Webサーバ(以下WWWサーバ)を中心とした情報提供網である。ところで、これまでのインターネットに関する議論は、そこでは、どんなことができるのか?とか、どんな風に利用するのかといったどちらかという情報利用の立場にたったものが多い。サーバ構築を解説する書物の多くも、HTML言語の基礎的な使い方に割かれていて本格的な情報提供の際に必要な技術について触れたものはほとんど見受けられない。また、情報提供サイトで独自に開発されてきた大部分の資源がWWWと協調しているとはいいがたいことも事実である。そこで、本文では、情報提供の立場から、我々がWWWサーバ構築にあたって用いた技術について報告する。まず、WDCMが抱えるデータについて触れ、これらのデータによるWDCM WWWサーバの構成の概略を示す。WWW利用の応用として、我々が実現したWWWサーバと既存のプログラム群(例えば、RDB)との融合について報告する。この技術の応用により、今までの資源を無駄にすることなく、ユーザにとってよりスムーズな情報抽出を支援するシステムへ変遷することが可能である。

WWWの導入とWDCMデータの特徴

World Data Center of Microorganisms(以下WDCM)は国際微生物株系統保存連盟(以下WFCC)のデータセンターとして活動している。我々が有するデータの中で代表的なものは、WFCCに加盟する世界各国の微生物系統保存機関から提供された、系統保存機関の全般情報(例えば、連絡先や事業内容といったもの)とその系統保存機関が保存している菌株名からなっている。現在オンラインで提供しているデータには、これ以外にも、「藻類の系統保存機関とその保存種」や「モノクローナル抗体に関する情報」等がある(図1)。さて、あるひとかたまりの情報を効率良く提供するためには、そこに含まれる情報の内容を十分に分析するとともに、その情報を利用する者がその情報を引き出す過程についても十分に検討しておく必要がある。後者について、我々のデータに当てはめて見ると、次の2つが前提となっている場合が多いと思われる。

- (1) ユーザには、興味ある系統保存機関が既にある、この機関の所在等を知りたい場合。
- (2) ユーザが興味ある微生物を特定していて、この微生物を手に入れることが可能な系統保存機関を知りたい場合。

すなわち、利用者にとって使いやすい提供形態は、最終的にいずれかの系統保存機関の情報にたどりつけるようになっていたことである。系統保存機関の情報はそれだけで大きな情報源である上にこの情報塊だけを検索する機会は極めて高いものである。そうすると、最も合理的なやり方は、すくなくとも、系統保存機関の一般情報は単一のデータベースとしてまとめて、微生物の学名等のその他の情報から切り放しておき、必要に応じて動的に保存機関の情報のデータベースのレコードを他のデータベース結びつけるようにすることである。これまでも、ハイパーテキストのような考え方を導入すれば、リンクと呼ばれる技術を使ってレコード同士を結び付けることは導入されていたが、これらの技術が実装されたシステムでも、リンクを設定するのにもかなりめんどろな手続きが必要であったし、静的なもので、あらかじめ設定されたものを途中で変更することも容易ではなかった。我々がWWWを導入した大きな要因は、WWWで導入されたHTML言語のリンク機能設定に着目したからである。現在、我々が提供しているシステムは図2のような構成になっている。代表的なデータ群はそれぞれ、WAISで管理しており、各レコードには必要に応じて系統機関の一般情報に関するデータベース(CCINFO)へのリンク情報を埋め込んだ(図3)。こうして、複数のデータベースで分散的に情報を管理するとともに、全体としてまとまりのよい構成が実現できたのである。

WDCMで管理している主なデータベース	
データベース名	内容
CCINFO	世界の微生物系統保存機関の一般情報
STRAIN	微生物の学名とその保存機関に関する情報
JFCC	国内の微生物の保存情報
ALGAE	藻類に関する保存情報
RCB	理研筑波センターに保存されている細胞情報
JTCA	国内に保存されている細胞の情報
HDB	モノクローナル抗体に関する情報
DSM	DSMに保存されている微生物の情報

注: WAISで管理されているものだけを挙てある

図 1.

```

<HEAD>
<TITLE>Bullera grandispora</TITLE>
</HEAD>
<BODY><PRE>
<H2>Bullera grandispora</H2>
Availability of Culture: E for Exchange, C for Charge, F for Free
<P>

    Bullera grandispora

<P>

<A HREF="http://database/CCINFO/216.html">IFO</A>(191): EC
</PRE></BODY>
    
```

図 3. LINK情報の埋め込み

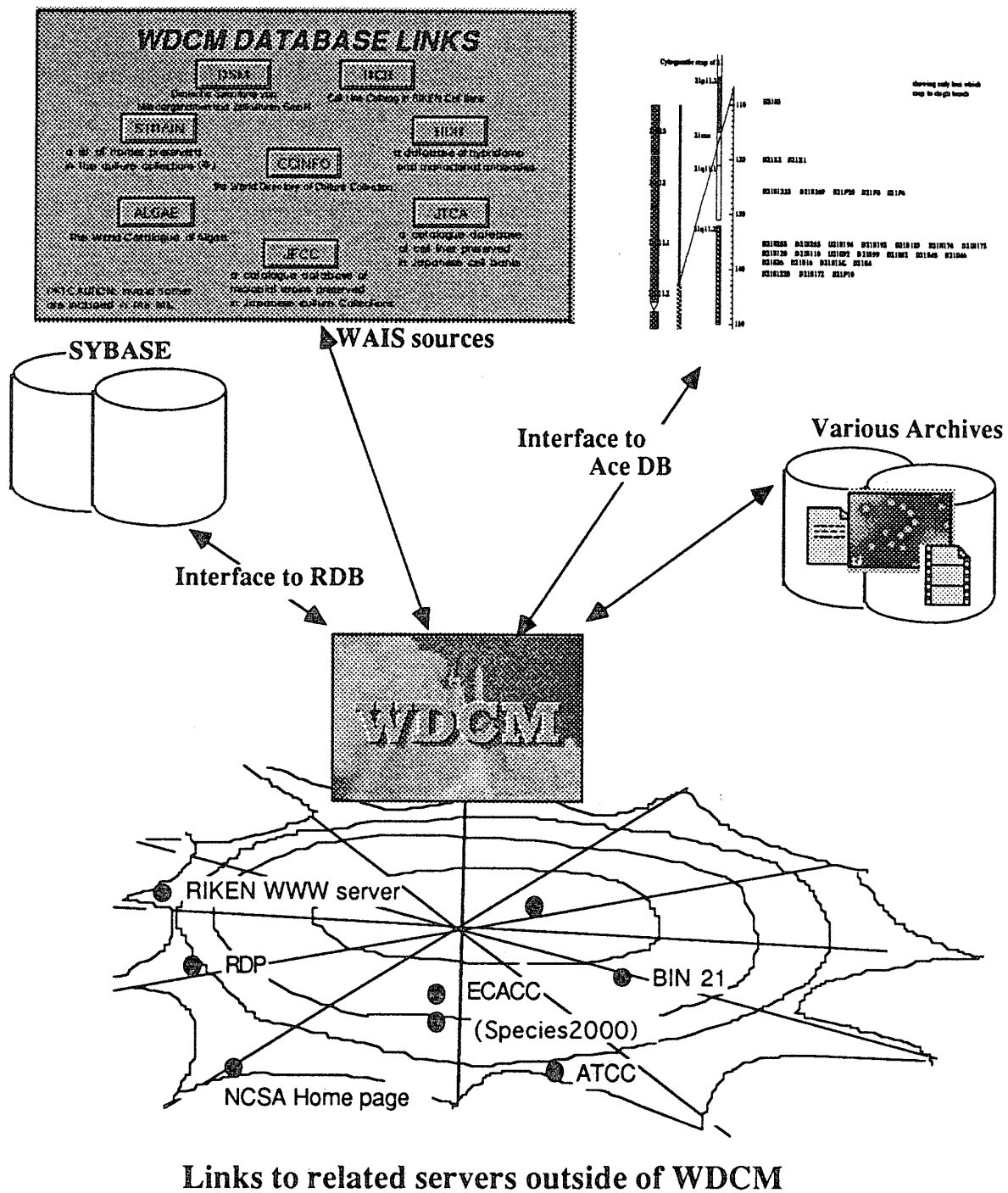


図2. WDCM Webサーバの構成

従来のデータ管理ツールとWWWの融合について

WWWで情報を提供する場合に、CERN httpd、NCSA httpdに限らず、バックグラウンドで動作する標準のデータベースシステムはWAISを用いることが、一般的であろう。しかし、かつてより情報を提供してきたサイトでは、独自のデータベースシステムを使用していたり、よく知られたRDB(例えばSYBASEやINFOMIX等)を使っていることもまれではない。また、WAISは元来フルテキストサーチ型のデータベースシステムであるため、データに数値が含まれていて、例えば、数値データの大小による検索はできない。あるいは、これから頻繁に取り込まれるであろうグラフィックデータの十分な管理は期待できないであろう。これらのことをうまく克服していくためには、まず、WAIS以外のDBとWWWの融合を考える必要がある。より、一般性を増すためには、他のDBシステムというよりは、WWW内から他のプログラムを制御し、そのプログラムからの結果をHTML形式に即座に変換し出力することを目標にすべきである。この技術が理解できれば、同様のやり方で、ほとんどのプログラムとWWWでリアルタイムでのデータのやり取りが実現できるはずである。我々が実践した一つの有用な方法は、CGI POST METHODを使ったやり方である。CGI POST METHODとは、HTML言語に予約された、`<FORM METHOD=POST ACTION=\ "http://.....">....</FORM>`タグを使って実現される。これにより、例えば、ユーザが動的に指定するデータベース検索のためのキーワード等も他のプログラムに引き渡すことが可能となる。例えば、我々のSYBASEとWWWのインターフェースの処理の流れは、

- 1) CGI POST METHODによりログイン名、パスワード、検索キーワードを得る。
- 2) 上記で得た値を引数としてisqlプログラムに渡し、isqlを使ってSYBASEを検索し、結果を得る。
- 3) 出力の形式としてContent typeをtext/htmlで指定する。
- 4) isqlによる検索結果をHTML形式に変換して標準出力に表示する。

となっている。CGI POST METHODと、Content typeの指定の組み合わせをうまく使うとWWWとは独立に作られた相当数のプログラム群を統一的に使うことが可能になる。前者によって引数の動的な指定が可能となり、後者の指定によっては、出力結果を見るためのMosaic以外のツール群の起動を制御できるからである。例えば、Content typeにsoundを指定し、標準出力に音声保存されたものを送れば、もし、受け手側で、音声出力用のプログラムを用意しておくと、このプログラムが連動し、ユーザ側には、しかるべき音声流れるといったことも容易に実現できてしまうのである。

まとめ

ここでは、WDCMが保有するデータを有効に引き出すためのシステムをWWWで

構築した事例を説明した。WDCMではCCINFOデータベースが一種のclearing houseとなり、利用ユーザにとって円滑なデータ抽出に貢献している。また、データ管理システムとして、WAIS以外のものを積極的に導入する方法の一例を解説した。POST METHODの使用により、WWWを通して既存のプログラムを制御する流れを確立し実装する手法を示した。

WWWの登場により、情報提供、利用の両者においてコンピュータネットワークの利用が活発化し、従来独立に提供されていた資源が統一的に使用できるようになってきた。しかし、WWWによる提供資源の急速な広がりには次の問題を引き起こそうとしている。WWWは、個人レベルで比較的簡易なハードウェアを用いて情報提供が行えるために情報提供サイトの数が爆発的に増えつつある。また、例えば遺伝子配列に関する情報等は、複数のサイトが独自に付加的な情報を加えて提供するようになってきているので、幾つかのサイト間でデータの整合性がとれなくなったり、データに精通しない利用者はやみくもにネットワーク上を渡り歩くだけで、結局有意義なデータにたどり着けないといったことが発生しつつある。こういった状況をそのまま放置すれば、ネットワーク上に相当の混乱が生じるであろう。そこで、次の課題として、複数の資源を統合した、いわば仮想分散環境を構築するとともに、ネットワーク全体を見渡して、各情報提供サイト全体で整合性のとれるような情報提供の仕方を考慮していく必要が生じてきていると思われる。

*理化学研究所

The Insutituton of physical and chemical resarch

** (株) メイテック

MEITEC co.