

WWW サーバの運用と問題点

総合情報処理センター

花田 英輔

E-mail: hanada@cc.nagasaki-u.ac.jp

1 WWW のサーバとクライアント

1.1 マルチメディアオンラインデータベース WWW

近年、「マルチメディアオンラインデータベース」が注目を浴びている。長い単語であるが、「マルチメディア」と「オンラインデータベース」の2つの部分から成る。

マルチメディアとは複数種類のメディア(媒体)を融合させたものであり、すなわち提供可能な情報として文字情報、静止画像情報、動画像情報、音声情報を統合的に扱うことができることを意味する。また、オンラインデータベースとはコンピュータネットワークを介して分散的に配置された情報を手元の端末からのコマンドで参照できることを意味する。

現在、このようなマルチメディアオンラインデータベースの代表として扱われているのが WWW (World-Wide Web) である¹。

WWW にはサーバとクライアントが存在する。WWW の場合、サーバとはデータを供給している部分をいい、クライアントとはデータを検索し表示する部分を言う。サーバとクライアントは同じマシン上にあっても、ネットワークで接続された異なるマシンに片方ずつあっても構わない。

なお、WWW におけるクライアントソフトウェアは、その表示部分が主な機能となるため、通常は「ブラウザ (browser)」²と呼ばれている。

1.2 クライアントとサーバの接続

クライアントとサーバの接続は、必ずクライアント側から行われる。接続の際にはマシン名とポート番号、参照したいデータの名前(ファイル名)が必要となる。WWW サーバが動作しているマシンで他のオンラインデータベースが動作している場合もあり、接続ポートを変えることで異なる通信手順を共存させることが可能である。

近年、WWW の普及と共に接続先の指定方法として URL (Uniform Resource Locator) がよく用いられるようになってきている。例えば総合情報処理センター(以下、本稿中では単に「センター」と呼ぶ。)のサーバを指定する URL は

<http://www.cc.nagasaki-u.ac.jp/>

¹Web は「くもの巣」という意味であるが、辞書名としておなじみの Webster の意味もあるようである。

²MSP 利用者にはおなじみであるが、ブラウズとは「表示する」という意味がある。

となる。これは、`www.cc.nagasaki-u.ac.jp` というマシンへの WWW 標準の手順である `http`(HyperText Transportasion Protcol) による接続を示している。なお、実際には最後にポート番号とファイル名を指定する必要があるが、ポート番号は予め標準番号を設定でき、またマシン名(もしくはポート名)の後ろに/だけをつけた指定方法でサーバにおいて予め指定したファイルを表示するようになっている。

本稿ではこういった WWW のうちサーバ部分について説明し、センターにおいて約 1 年間運用してきたその実績や問題点について紹介する。

WWW のサーバは、マシン上で動作しているプロセス名やソフトウェアの名前としては `httpd`(HyperText Transportasion Protcol Daemon) ということが多いが、本稿では WWW サーバ(または単にサーバ)と呼ぶことにする。

2 WWW サーバソフトウェアの入手

WWW サーバはソフトウェアである。

WWW サーバソフトウェアは、ワークステーション用としては NCSA³(National Center for Supercomputing Applications の略)、CERN⁴(Conseil European pour la Recherche Nucleaire の略) の 2ヶ所から発表されているフリーウェアのものがよく知られ、広く用いられている⁵。

また、サーバソフトウェアは Macintosh や Windows 用のものも発表されており、その点ではパーソナルコンピュータであってもサーバの運用ができる。

いずれも、`anonymous ftp`⁶を用いて入手可能である。WWW ブラウザを用いてそれぞれのサーバに接続して入手することも可能⁷である。

入手後は、展開、設定、必要に応じてコンパイルなど行っただけでテスト運用を行うこととなるが、コンパイル前やテスト運用中には 3章に述べるような項目についての検討を行う必要がある。

3 WWW サーバの機能と設定

以下、ワークステーション用のサーバについて、その機能と設定方法について説明する。ここでは主に 2章であげた NCSA と CERN から発表されているものを基準として述べる。

3.1 サーバが持つ機能

WWW サーバの基本は情報の保持と提示であるが、これ以外にも様々な機能が必要である。ここではこれら機能について説明する。

3.1.1 データの提示

WWW のサーバとしてまず持つべきデータ提供機能であるが、その基本単位は「ページ」という。

³アメリカのイリノイ大学にあるソフトウェア開発チーム。

⁴スイスとフランスの境界上にある欧州素粒子物理研究所。WWW を開発したのはここである。

⁵ワークステーション用のサーバソフトはこれら以外にも数種類ある。

⁶匿名 `ftp`。誰でも利用できるファイル転送。

⁷ブラウザを用いた入手では、圧縮されているファイルが自動的に展開される場合があるので注意が必要。

ページには文字情報と静止画像情報を同時に提示する機能がある。また、ページ内の情報には他の情報や他のサーバの情報への入口⁸を設定することが可能となっている。

ページデータの設定方法や文法は他稿 [1][2] に譲るとして、これ以外にもサーバが有すべき基本的な機能として

- ファイル名を指定しない場合の入口ページ (「ホームページ」という) の指定
- ポート番号を指定していない場合の標準ポート⁹の指定
- テキスト形式データの文書としての提供
- 画像データ (静止画像/動画像) の提供
- 音声データの提供

がある。

しかし、実際にはこれだけではなく次のような機能も提供可能である。

- gopher や WAIS といった他種類のオンラインデータベース機能との連携
- ftp 等で用いるためのファイル検索システム archie 機能の提供
- UNIX コマンドやプログラム、シェルの実行と結果表示
- 電子ニュースの表示
- 予め設定したアドレスへの電子メールの発信
- サーバ上へのデータ追加/変更

ただし以上のうち電子ニュース/メールについては、現在のところ主にブラウザ側の問題があり、日本語データの取り扱いができない場合が多い。

日本語データの問題については4章で述べる。

3.1.2 データ管理機能

サーバを運用するマシンが専用マシンでない限り、運用しているマシンには様々なデータや一般ユーザデータなどが存在することになる。また、WWWサーバに置くデータそのものについても、範囲を限定して公開すべきデータがある。

そこで、サーバの機能としてアクセス権の設定と実行が必要となる。その方法はソフトウェア毎に異なるので、ここでは前述の NCSA および CERN のサーバソフトウェアを例にとって機能を説明する。

まず、アクセスの可否はブラウザから送られてくるデータ要求中にあるアクセス元 IP アドレスを用いて判定可能である。

アクセス権は IP アドレス1つ1つについて設定することも可能であるが、サブネット単位、ドメイン/サブドメイン単位での設定も可能である。

⁸これをポインタと呼ぶが、情報を結び付けることを「リンク」という。

⁹通常は 80 にすることが多いようである。ただし、このポートはマシンの root でないと使用できない。

また、CERN のサーバを用いる場合はユーザ ID とパスワードを設定することも可能¹⁰である。

アクセス権の内容は、読み取り権、変更権といった UNIX でのアクセス権と同じ様な内容である。

データ配置から見た場合、ディレクトリ単位でアクセスに関する情報ファイルを置くことでアクセス権を変更できる。即ち上記の各設定はディレクトリ単位で可能である。

これにより、アクセス権毎にデータ格納ディレクトリを変更することで、多様なアクセス権設定が可能となる。

実際の設定は、設定用に予め用意されている configuration ファイル¹¹に書き込むか、該当ディレクトリに設定ファイル¹²をおくことで実行される。何も設定しなければデータは自由に参照される。configuration ファイルと設定ファイルの優先度は configuration ファイルにおいて設定可能である。

設定ファイルの文法については、それぞれのドキュメントファイルをご覧いただきたい。筆者の感覚としては NCSA のサーバの方が設定方法は簡単である。

3.1.3 参照記録の保存と分析

サーバを運用する場合、その参照記録は保存の上統計データとして活用することでデータの更新や追加の際の資料とすることができる。

WWW サーバソフトウェアでは NCSA、CERN のどちらにも参照記録をログとして保存できる。保存形式は CERN のサーバで NCSA 側に合わせる事が可能で、かつ NCSA 側の方がわかりやすい¹³。

NCSA サーバのアクセスログの例

```
133.45.xx.yy user - [05/Jan/1995:10:50:55+0900] "GET /ISC/ISC.gif HTTP/1.0" 200 69147
```

ログ中の情報は、次のような順序である。

1. サーバが直接参照されたクライアントが存在するホストの IP アドレスもしくはホスト名
2. 参照しているユーザ名 (ただし、pidentd という別のサーバを起動した場合のみ表示される。起動していない場合は - となる。)
3. 要求を受け付けた日付と時刻
4. 要求内容
5. 結果コード
6. 転送バイト数

¹⁰センターのサーバでは実施例はない。

¹¹NCSA のサーバでは access.conf, CERN のサーバでは all.conf もしくは prot.conf

¹²NCSA のサーバでは htaccess, CERN のサーバでは .www.acl

¹³と、筆者は考えている。

結果コードには次のような意味がある。

内容	意味
200	正 常
302	Redirected Requests
400	Bad Requests
401	Unauthorized Requests
403	アクセス権により拒否
404	要求されたデータ無し
500	サーバ側の異常
501	要求自体 (内容) の異常

貯蔵したログは活用するべきであるが、サーバソフトウェアには保存したログの分析機能が無い。そこで、分析のためのツールを用いることになる。センターのサーバ運用においては、幸いなことに電子ニュースに投稿されていた、perl で書かれた2つ¹⁴のツール WebReport と wwwstat¹⁵を用いることで統計資料を作成することが容易に可能となった。

3.1.4 データ提供機能の追加部分について

3.1.1の項で述べた追加機能は、cgi というプログラム機能を用いることで提供可能となる。

cgi は Common Gateway Interface の略である。実体は C 言語で書かれたプログラムであったり、UNIX のシェルや perl などの言語を用いて作成できる実行可能なコマンド、またはコマンド群である。そのいくつかは CERN や NCSA の WWW サーバから入手可能である。もちろん自作も可能である。

ただし、同じ cgi であってもサーバの種類が異なると動作しないものもあるので、十分な注意が必要である。

もし cgi を自作する場合は、その実行権と cgi を置くディレクトリについて注意しなければならない。間違った場合は、cgi の内容が正しくても実行できない。

4 サーバ運用上の問題点

センターでは以上のような設定などを行った上でサーバを運用しているが、センターでサーバを運用開始するに当たって問題となった点について述べる。以下にあげる問題点はどこでサーバを運用するにおいても問題となるはずである。

1. サーバの公開範囲

現在、WWW サーバの運用開始を宣伝する場合は例えば電子ニュース等¹⁶での広報があるが、それ以外にも国内のサーバを一括管理している NTT のサーバに登録すれば、NTT の「国内のサーバ一覧」に掲載され、組織外からの参照が容易となる。なお、九州内では KARRN(九州地域研究ネットワーク) のサーバに登録すれば NTT のサーバに登録されたことになる。

しかし、組織外からの参照を許すと言うことは、通常全世界からの参照を許すことになるので、当然のことながら組織で守るべき秘密について検討が必要となる。

¹⁴他にもあるが、ここでは紹介しない。

¹⁵wwwstat はセンターの WWW を通してダウンロード可能である。

¹⁶日本では fj.net.infosystems.announce や fj.net.infosystems.www といったニュースグループで告知や議論がなされている。

そこで、サーバ自体の公開範囲とデータ毎の公開範囲について検討の上、サーバ自体については登録するか否かを決定し、またデータ(ページ)単位ではデータを置くディレクトリを分けた上で個々のディレクトリについてアクセス権の設定を行わなくてはならない。

2. データディレクトリとデータ形態、内容

データのアクセス権との関係でデータディレクトリを分割する必要があることは上に述べたが、それ以外にも次のような問題もある。

- データをどのファイルシステムにあるディレクトリに置くか
- ユーザディレクトリに各ユーザが自分のデータを自由に置くことができるようにするか
- データディレクトリ自体の(マシン上の)ユーザアクセス権の設定

これらについては、予め設計を行ってからデータ配置しないと、公開後のデータ再編成は非常にむづかしい。

また、データ形態ではとくに画像ファイルに関してその形式によって同じ画像でも占有データ量が異なるので注意が必要である。静止画像では TIF 形式が非常に大きなデータ量を要し、現状では JPEG 形式が最も小さくて済むようである。

さらにデータの内容の面では、既に述べた機密の問題以外に

- 情報は正しいか
- 不必要なデータを含んでいないか
- 引用データ等が著作権を侵害していないか
- 画像情報が著作権や肖像権を侵害していないか
- 個人のプライバシー情報に関わっていないか
- 公務員や公的機関が運営する場合、特定の企業等を利する情報を含んでいないか
- 情報が変化した場合に修正(メンテナンス)する要員を確保できているか

等の様々な問題が有る。

これらを考えた場合、安易な(特に個人的な)サーバ立ち上げは慎むべきかもしれない。

3. 標準設定

ネットワーク上の問題として、ブラウザが指定するサーバの名前とポートの問題、そして標準の入口ページ(「ホームページ」)の設定が必要となる。これらは概ね 3.1.2 で述べた configuration ファイルで設定する。

まず、サーバを運用するマシンのホスト名を設定しなくてはならない。ホスト名は固有の名前でもよいが、マシンが変更となる可能性も考えると固有の名前とは異なるものをつけた方がよいと考えられる。

通常、サーバを起動するマシンがサブドメイン (例えば xxx.yy.ac.jp というサブドメインを例にする) の中に 1 台しかない場合は、

`www.xxx.yy.ac.jp`

と言ったようにホスト固有名の部分を `www` と置き換えた名前をつけ、ネームサーバ¹⁷で管理するケースが多いようである。

次に、ポートは試験的運用の場合等マシンの root 権限を用いず起動するためにはその番号が 1024 以上でなければならないが、root 権限を用いた運用では 80 を用いることが多いようである¹⁸。なおこの点については、サーバソフトウェアの起動方法によって設定を行うファイルが変わるので、設定時に十分注意する必要がある。

最後に、ホームページ名は通常 `index.html` とすることが多いが、このファイルをどのディレクトリにおくかを設定する必要がある。ホームページをおくディレクトリはデータディレクトリの中でも最も上位にある必要があるが、このディレクトリを `ServerRoot` と呼んでいる。いずれも `configuration` ファイルで指定する。

4. 漢字コード

日本語を用いたページやデータを置く場合は、予め漢字コードについて検討する必要がある。

これまでは、UNIX の X ウィンドウ上で動作する Mosaic¹⁹を用いて参照するケースが多く、Mosaic が余分な操作無く表示できる jis コードを選択するケースが多かった。

しかし、近年 Macintosh や Windows マシンがネットワークに IP 接続される数が急激に増えており、かつその場合に利用できるブラウザは、現状²⁰では、Macintosh では SJIS もしくは EUC、Windows では SJIS、NeXT では EUC で書かれた日本語データを表示できるブラウザがあることが確認されているのみで、いずれも jis コードで書かれた文字を表示することができない。そこで、データ提供側で対処する必要がある。

1 つの方法として、作成した文字データの漢字コード変換を行って 3 種類用意することも可能である。しかし、当然のことながらメンテナンス量が増えるためあまり現実的ではない。

そこで、漢字コードを変換するツールを間にかませる方法を検討しなければならないが、現在のところ `delegate` というツールがある。

`delegate` は電子技術総合研究所 (電総研) の佐藤氏が開発し、現在もバージョンアップが続いている。入手は `anonymous ftp` を用いて可能であり、フリーウェアである。よく用いられる機能としては proxy 機能²¹、cache 機能²²、漢字コード変換機能がある。

そこで、このうち漢字コード変換機能を活かしてユーザからの接続先を `delegate` にしてもらうことで漢字コードの面の問題を解消することができる。

¹⁷ コンピュータネットワーク上のホスト名と IP アドレスの対応表とを考えてください。

¹⁸ root 権限以外で WWW サーバを起動する場合は 8001 や 8080 といったポートがよく用いられている。

¹⁹ ほとんどの場合、NCSA XMosaic の多言語パッチ版

²⁰ フリーウェアのものに限って調べた。

²¹ proxy とは代理のことで、参照者を代行する機能と考えればよい。会社などで組織外とのデータ流通に制限がある場合に役立つ。

²² 一旦参照したデータを保存しておく機能。ネットワーク上のデータ流量を押えることができる。

ところが、この方法には1つ問題点がある。3.1.3の項で述べたように、特にセンターと言った共用的な組織でサーバを運用する場合は、参照記録を保存し、解析する必要があると考えられるが、delegateを通してサーバにアクセスした場合は、サーバ側に残る記録はdelegateが動作しているマシンからのアクセスとして残ってしまうのである。delegate自身にもアクセス記録をとる機能はあるが、delegateは漢字コード毎に起動しなければならない²³ので、接続先毎に記録を集計したうえで統計にかけるなどの操作が必要である。記録ファイルの変更などのタイミングの上でかなりの検討が必要であり、またdelegateを経由せずに直接サーバへ接続する場合を許すと、サーバの記録との兼ね合いがあるためなかなか難しい。

5. アクセス記録(ログ)の処理

アクセスの記録が残ることは既に述べたが、一定の周期でこれを処理し、また一定期間の後に過去の記録を破棄しないとディスク容量の点で限界がきてしまう。そこで、一定周期(週/月)単位で記録を集計し、また記録ファイルから過去の記録を消去しなければならない。

集計については3.1.3で述べたようなツールを利用すればよいが、過去の記録を順次選択して消去してゆくのは煩わしく、かつ記録ファイルを見るのは通常管理者だけなので、記録ファイルをまるごと消去もしくは取り替えてしまうのが簡単である。ただし、過去の記録を直ちに消去するのではなく、一定期間過去の分も世代管理しつつ残す必要があると考えられる。

5 総合情報処理センターにおけるサーバ運用

本章では、センターにおいて実験的に運用してきたサーバについて、その運用形態やアクセス実績について述べる。

5.1 サーバ立ち上げの経緯と当初の設定

センターのサーバは、もともと平成6年3月14日に長崎厚生年金会館において開催されたKARRN協会の講演会におけるデモンストレーション用として作成された。

その際、大学やセンターの紹介だけではサーバアクセスの際に面白みがなく、デモンストレーションとしての見栄えにも問題があるということで、長崎市の観光案内をデータ化²⁴した。

また、講演会終了後もセンターにおけるオンラインデータベースの実験及びデモンストレーション用として利用できることを目的としたので、主に学内からのアクセスを考えたデータ配置や入力を考えている。

その他、講演会には県外からの参加者が大変多かったため、帰りの時刻表が欲しいと言う要望が講演者から出され²⁵、長崎駅、浦上駅発のJR、長崎の各ターミナルからの長距離高速バス、長崎空港からの飛行機の時刻表を掲載することになった。

²³接続ポートを変えておけば並立可能である。

²⁴データ化といっても、大学紹介も含めて筆者が1人ですべて原稿から作成し、必要な画像データも揃えた。最初のデータは1週間で準備した。

²⁵というよりも、作者=筆者の趣味によるところも大きい。

なおデータ掲載に当たっては、長崎大学事務局総務部に問い合わせた結果学内の情報(キャンパス配置図など)は講演会当日のみの公開許可となった。また観光案内については長崎市観光協会発行のパンフレットに掲載されている写真を用いたが、当初了解をとる時間がなく同年3月25日に事後承諾を得ている。時刻表データは担当である筆者がその時点で最新の時刻を入力し、現在も時刻変更の度²⁶にデータ入力を行っている。

利用したサーバソフトは、立ち上げ当時 CERN のサーバソフトがバージョン 3.0 のプレリリース²⁷途中であったこともあり、設定も容易な NCSA のサーバソフトを用いている。現在は別のマシンで両方のサーバソフトを用いてテスト中であるが、様々な運用上の問題点が未解決であり、かつそれに取り組む余裕がない。将来的には CERN のサーバソフトに移行する希望を持っている。

5.2 データの現状

このように各データを入力して講演会に臨んだが、終了後の継続的実験運用の中で、時刻表データの更新以外に多少データの充実を図っている。その結果現在(平成7年1月現在)次のようなデータ(ページ)を用意している。(*マークがついているものは講演会終了後新たに追加したデータ)

- メインページ(大学全体のメインページを兼ねている)
- 長崎市紹介ページ「ようこそ長崎へ！」
- 個々の観光地ページ
 - 路面電車*
 - 日本 26 聖人殉教記念碑
 - 大浦天主堂
 - グラバー園
 - オランダ坂
 - 東山手洋館群*
 - 崇福寺(中国寺)
 - めがね橋
 - 思案橋
 - 諏訪神社*
 - 出島
 - 新地中華街
 - 平和公園
 - 浦上天主堂
 - 原爆落下中心地
 - 如己堂*

²⁶ 飛行機のダイヤはほぼ毎月変るが、他のダイヤは年に1回変るか変らないかである。

²⁷ 正式提供の前にバグ調査などの目的で試験的に提供すること。

- 稲佐山
- 長崎水族館*
- 長崎の夜景 (拡大画像データあり)
- 長崎の祭
 - ペーロン競漕
 - 精霊流し
 - 長崎くんち
- 長崎発の時刻表 (JR, 飛行機, 高速バス)
- 長崎空港との交通手段*
- 総合情報処理センター紹介ページ
 - 総合情報処理センターシステム構成図 (画像データ)
 - NUNET 接続概念図 (画像データ)
 - アクセス記録 (毎週月曜日の深夜に自動更新)*
- NUNET ニュース (各号)*
- 長崎大学の紹介 (いずれも学内からのみアクセス可)
 - 本学の沿革
 - キャンパス毎の建物配置図 (画像データ)
 - 本学の組織図 (画像データ)
 - キャンパス配置図 (画像データ)
- WWW 紹介のページ (学内からのみアクセス可)*
 - ページの例 (浦上天主堂のページで兼用)
 - 音声データの例
 - 静止画像データの例 (稲佐山からの拡大夜景で兼用)
 - 動画データの例

データ形態を分類すると, html(HyperText Markup Language) 形式のページが 96, テキスト形式のものが 12, 静止画像はページに張り付けたものも含めて 58 枚, 動画は 3, 音声データも 4 用意している²⁸. なお, 動画と音声データはいずれも WWW 紹介用のみである.

データ量で見た場合, 全体で 14MB あるが, 動画や音声データは各ファイル毎のデータ量が大きいので, これら以外のデータでは 7MB 程度となっている.

データのリンク構成を概念的に図示すると, 図 1 のようになる. 図中各文字がページ内容を示すが, 図の上の方にあるページを上位ページ, 下の方にあるものを下位ページと呼ぶ. ただし下位のページを持つページは, その話題に関するホームページと見なすことができる.

²⁸現在テキスト形式のものを html 形式にする努力中である.

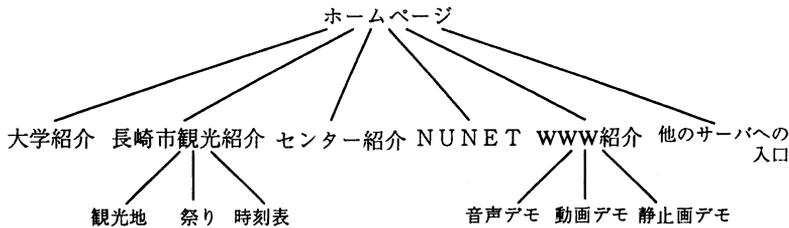


図 1 センターの WWW サーバにおけるデータ構成

各ページのアクセス権は、データディレクトリを観光地データ、センター紹介データ、大学紹介データ、NUNET、WWW 紹介 (デモ) の 5 つにわけ、そのうち大学紹介データと WWW 紹介データのディレクトリについては、アクセス要求があった場合に要求元のマシンの IP アドレスが 133.45.*.* という IP アドレスであるかどうかを調べる²⁹ような設定にしている。また、個々のユーザディレクトリにあるデータはすべてのマシンに対して公開していない。cgi についても、サーバをおいているディレクトリでの実行しかできないように設定している。

各ページの構造は、ホームページに当たるもの、具体的には全体のホームページと観光地紹介のホームページには画像は使用していない。また、観光地データなど末端に当たるページに関しては前述の観光パンフレット上の写真³⁰をカラーキャナーで取り込んだものを 1 枚と説明文、観光地データの場合は交通手段をできるだけ簡潔に入れている。これらはデータ転送時間を可能な限り短くして見やすいサーバにするためである。

また、時刻表に関しては html 形式にした場合に現在のところブラウザで表形式での表示ができないためテキストデータとしたが、html 形式化するべく調査中である。

各ページは、WWW の紹介と NUNET News をページを除いてすべて日本語版と英語版を用意し、かつ日本語データについては学内で WS よりもパーソナルコンピュータの方が多いいことを考えて Shift-JIS と EUC の 2 種類のみ³¹とした。

5.3 参照記録の利用

3.1.3でも述べたように、運用上参照記録を分析することは、参照傾向を掴み、内容の充実等に活かすために必要である。これは実験的に運用している場合でも同じで、むしろ実験的運用であるためにより細かい分析が必要となる。

センターのサーバでは、毎週 1 回ログファイルを更新し、更新直後のファイルについて解析している。即ち、1 週間分の記録を毎週分析している。古いログファイルは 5 世代残している。即ち最大で 5 週間前の参照記録については個々のログが残っていることになる。ログファイルの大きさは 50KB 程度から 300KB 以上になることもある。

ファイルの更新と解析ツールの起動はワークステーション上でタイマーにより自動起動する。現在は毎週月曜日の深夜 (23:00) に行っている。この時間であれば、アメリカは月曜日の早朝、ヨーロッパは月曜日の夜であり、おそらく利用者が少ないと考えた結果である。

²⁹ 長崎大学はネットワーク的には B クラスの IP アドレスを交付されており、すべて 133.45 で始まる。

³⁰ 路面電車など不足分は筆者が撮影した写真を使用した。

³¹ ワークステーションでもこの 2 つのコード体系でのテキストは読むことができるので、心配はない。

解析ツールとしては、3.1.3で紹介した WebReport と wwwstat を両方用いている³²。

WebReport は出力がテキスト形式であるのに対して wwwstat は html 形式で出力する。内容的にはほぼ同じで、アクセス回数とアクセス量を全体、日別、時間帯別、アクセス元別、データファイル別にそれぞれまとめる。

解析結果は担当者が見るのは当然として、それ以外に wwwstat の出力をページデータとしてセンター紹介ディレクトリにおいている。後述するが、このページの参照もかなりある。サーバ上に置いている解析結果は最新のもののみとしているが、担当者用の解析結果は全結果を蓄積している。

5.4 運用実績

サーバのアクセス状況について、1994年4月から年末までの記録を元に分析してみる。

5.4.1 アクセス量

まず月別アクセス状況を表1にまとめる。この数値は、すべてのアクセス数と送出力である。

表1 総合情報処理センターのWWWへのアクセス状況(アクセス件数とデータ量)

月	4	5	6	7	8	9	10	11	12
アクセス件数	346	531	1439	2823	4689	2094	1684	5568	1480
送出力(KB)	6648	24528	33741	72069	123385	434235	38887	217765	359214

これによると、7月に利用の山があり、また11月はとび抜けて多い。これは、全学教育の一般情報処理科目における参照の時期に当たっていると考えられる³³。今後も授業においてコンピュータとネットワークの融合を説明する際に利用されることが予想される。

またアクセス件数1回当たりの平均送出力データ量は27.6KBである。センターのホームページの大きさは約3KB、観光地データの場合は文字データが600B程度、画像データが50KB程度であるので、ホームページだけでやめた場合や間違いを含んだ要求³⁴が多いと言うことが考えられる。

ブラウザには文字データのみ取得し、画像データはコマンドにより別途必要な時のみ取得するよう機能もあきるが、記録を見る限りそういう方法をとっているユーザは少ない。

また、個々のアクセスデータを見た限りでは、アクセスした場合にホームページだけでやめてしまう場合と、一通り観光地データを見る場合が目だった。従って、最初のアクセス程度でやめるユーザの方が多かったと言うことになる。やめてしまう原因としては、

- 日本語データをSJISとEUCだけにしたこと
- 内容に問題がある
- 学外のサーバへの入口として利用されている

といったことが考えられる。

³² WebReport については若干改造が必要であった。

³³ 少なくとも11月には筆者が担当している全学教育の情報処理演習において、合計で少なくとも200回のアクセスを行っている。

³⁴ 間違いを含んだ要求にはエラーメッセージを返すが、その大きさは300B~500Bである。

5.4.2 アクセス元

アクセス元の分類では当初圧倒的に国内(学内)が多いと考えた。実際学内からのアクセスが圧倒的に多いが、4章に述べたようにNTTに登録していないにも関わらず外国からのアクセスも多い。

表2に国別のアクセスサイト³⁵数をまとめる。

表2 総合情報処理センターのWWWへのアクセス状況(国別)

サイト数	国名	サイト数	国名
171	日本	3	フランス
145	アメリカ	2	スイス, フィンランド
9	カナダ	1	イタリア, ノルウェー, オランダ,
8	イギリス		ギリシャ, マレーシア, 韓国,
6	ドイツ		シンガポール, アルゼンチン,
5	スウェーデン		イスラエル, スロバキア
4	オーストラリア	合計	20ヶ国, 365サイト

表2でわかるように、アメリカ以外はまだ少ない。

アメリカ以外で特徴的なところでは、イスラエルがHigh School, その他は大学が多い。これらの国々では一般企業へのネットワーク(WWW)の普及が進んでおらず、学校の情報系の学部から参照されているものと考えられる。

逆に、アメリカでは大学やコンピュータメーカー以外にNASAや陸/海/空軍、州政府および赤十字社、警備会社といったところからの参照があった。

センターのサーバがいずれにも登録されていないにも関わらずこれだけの参照を受けると言うことは、長崎という地名が世界的に有名であることも考えられるが、知らないうちにどこかのサーバに登録(リンク)されているということも考えられる。今のところ国外に関してそのような事実は掴んでいない。

5.4.3 ページ単位

各ページ(話題)毎のアクセス状況を表3および表4にまとめる。なお、アクセスの際に間違ったコマンドや存在しないデータを要求していた場合があるので、これらは除いた。

これらの表からわかるように、コード形態ではEUC版の参照がかなり少ない。また、各ページのうちホームページに直接リンクされているページのアクセスが多く、また個々の観光地ページは観光地ホームページの掲載順の上の方から多いという傾向がある。

このことは、ページおよびページ構成作成の際に、アクセスして現れる画面上に見えるか否かがアクセス頻度を左右することを意味していると考えられる。またこのことは、特定の意識や目的をもってアクセスしないユーザの傾向を示していると言え、今後ユーザを引き付ける³⁶ためのページ作成方法の方向を示したものと言える。

5.4.4 参照の傾向とサーバとしての対策

以上の結果と運用および他のサーバを参照してきた経験から、次のようなことがわかる。

³⁵サイトとは、この場合大学、企業、政府機関など、ネットワーク上の組織のこと。

³⁶サーバ運営の方針に寄るが。

表 3 ページ毎のアクセス回数(その1)

ページ名	英語版	SJIS 版	EUC 版
長崎市紹介ページ	906	810	95
個々の観光地ページ			
路面電車	192	243	55
日本 26 聖人殉教記念碑	193	205	24
大浦天主堂	204	197	33
グラバー園	164	202	27
オランダ坂	109	137	18
東山手洋館群	53	67	6
崇福寺(中国寺)	40	60	3
めがね橋	65	123	11
思案橋	41	74	9
諏訪神社	32	57	6
出島	47	82	13
新地中華街	40	95	12
平和公園	53	75	11
浦上天主堂	34	81	12
原爆落下中心地	50	70	12
如己堂	8	36	1
稲佐山	53	82	10
長崎水族館	44	87	9
長崎の夜景	71	173	19
夜景の拡大写真	風頭 84	稲佐 211	
ペーロン競漕	25	57	4
ペーロン拡大写真		4	
精霊流し	30	81	12
精霊流し拡大写真		6	
長崎くんち	54	117	11
長崎くんち拡大写真		24	

- サーバは公開してもしなくても運用していれば参照しに来る。
- 参照されたいページは上位ページの最初の方、1画面に収まる範囲内にいれるべきである。
- ホームページはわかりやすく、できるだけ短く作るべきである。
- ホームページには画像データを入れるべきでない。入れるとしても最小限にすべきである³⁷。

特にサーバ名だけを指定した参照要求で送り出すホームページはできるだけ簡単かつ魅力あるものにすべきである。

また、特によく参照されるページやページ群がある場合には、それらをホームページから直接参照できるようにするのがよい。その最もよい例は国立がんセンターのWWWホームページから参照できる気象衛星「ひまわり」の画像データである。このデータは gopher

³⁷データ転送に時間を要し、参照する意欲を減少させる。

表 4 ページ毎のアクセス回数 (その 2)

ページ名	英語版	SJIS 版	EUC 版
長崎発の時刻表			
JR	57	90	6
飛行機 (各月の合計)	44	52	8
高速バス	32	71	7
空港との交通手段	'95 年 1 月から掲載		
センターの紹介			
センター紹介ページ	458	233	81
センター拡大写真	81		
システム構成図	116		
NUNET 接続概念図	541		
サーバアクセス記録	122		
長崎大学の紹介			
本学の沿革	164	98	15
キャンパス毎の建物配置図 (画像データ)			
文教キャンパス	77	日本語版 62	
坂本キャンパス	36	日本語版 47	
片淵キャンパス	19	日本語版 16	
本学の組織図	68	日本語版 30	
キャンパス配置図	27	日本語版 18	
NUNET ニュース			
創刊号	-	42	6
第 2 号	-	10	13
WWW 紹介 (デモ)			
WWW 紹介のページ	-	94	6
音声デモ	86		
動画デモ	63		

の中に入っているが、gopher から参照しようとする と接続初期画面からメニューを 4 回ほど選択しなければならない。これを、WWW サーバのホームページから直接指定可能としてある。

5.5 センターサーバの今後について

センターのサーバは、実験的な運用を続けている。これは 5.1 にも述べたように学内のデータに関して公開可能かどうかの判断ができないまま運用してきた点と、センターが全世界向けのデータベースを運用することが業務として認められるかどうかの判断が下されていなかったことによる。

今後は、まず公式に全世界に公開するかどうかをまず決定³⁸していただき、その後内容の充実をして行く必要があると考えられる。

充実すべき内容としては、学内の情報公開という点で学部紹介であったり、例えばネットワーク利用無償 ID のアドレス帳といったサービス³⁹も考えられる。

ただし、これら考えられるサービスをすべて総合情報処理センターで行うには問題点も

³⁸ 現在は、その決定を下す機関も不明確である。

³⁹ WHOIS というデータベースとの連携により可能となる。すでにサービスしている大学もある。

多い。すなわち、メンテナンスやデータ入手などを一手に引き受けることは他の業務に支障を招く原因になる。かといって、すべての部局にサーバを上げる必要があるかといえば、これも技術的／手続き的に無理があると思われ、これからの情報公開の体制作りには任された問題点と言える。

6 まとめ

本稿では、WWWサーバを運用する上で技術的、体制的に検討すべき点、問題点と実際に運用している総合情報処理センターの例、及び運用実績について述べた。

現状は実験的なサーバをほぼボランティアの形で運用しているが、今後は情報公開の立場から全学的な検討がなされ、認知された運用とそれに必要な体制が整えられることを希望している。

そのためには、「マルチメディア」や「WWW」、「Mosaic」といった言葉や、その実体について広く周知し、簡単に利用できることを知らしめなくてはならないと考えている。

最後に、本稿がセンター以外の部局でのWWWサーバ立ち上げの際に参考となり、本学が全学的に開かれた、かつネットワークを活用した先進的な大学となることを望みます。

参考文献

- [1] 野崎，花田，馬場：「インターネットによるカラー画像情報発信」
長崎大学総合情報処理センター「センターレポート」第14号,1995
- [2] 馬場，花田，野崎：「HTML文書作成入門」
長崎大学総合情報処理センター「センターレポート」第14号,1995