

## 第 14 部

# WWW による情報提供



# 第 1 章

## NIR-TF と、その背景

過去の UUCP によるコネクティビティ、そして、かなり一般的にインターネットが使うことが出来るようになった時点においても、インターネットのアプリケーションの核はメールとニュース、そして ftp などによるファイル転送であったと言えよう。また、telnet 形式でアクセス出来るシステム(たとえば、図書館情報検索など)は、いろいろな形式で提供されていたし、古来の端末形式のアクセスが生き残る限り、今後も提供されるであろう。

しかし、本来、インターネットの利用をそれにとどめていては、宝の持ち腐れとも言える。インターネットの特質を生かすことは、インターネットを構築するのと同様に重要である。インターネット、つまり、WIDE プロジェクトの名称の由来である、広域分散環境におけるアプリケーションというものがユーザが利用することが出来て初めて有効利用されていることになる。

広域分散環境の利用形態の一つとしては、ネットワーク化された情報の提供や入手であろう。これを Networked Information Retrieval(以下 NIR と略)と言う。

ところで、すこし前には、「インターネットで何ができるか？」との問いに対して、電子メール、電子ニュース程度しか思いつかなかったともいえる。

ここ数年で、インターネットが急速に整備され、高速データ通信が可能になるにつれ、マルチキャスト通信を用いた会議システムである「vat(Visual Audio Tool)」および「nv(NetVedeo)」のようなシステムの運用が可能になり、ようやく、インターネットにおけるマルチメディアが日の目をみることとなった。とくに、CERN が開発した WWW(World Wide Web) システムは、そのビジュアルな点から各方面で話題を読んだ。社会に与えた影響も大きかったように思える。このシステムの登場が、全世界的な、いわゆる「インターネット・ブーム」を引き起したといっても過言ではないであろう。

WIDE プロジェクトにおいては、本報告書に紹介されているように、多くのワーキンググループがある。その中には、マルチメディアワーキンググループや ISODE ワーキンググループのように、アプリケーションを主眼としたワーキンググループがある。また、WIDE project のマルチメディアワーキンググループでは、昨年度から WWW システムの調査と評価を行っている。しかし、残念ながら WWW やそれに類したアプリケーションを取り扱

うのに適したワーキンググループは無かった。<sup>1</sup>

加えて、WIDE プロジェクトは、雑誌媒体などを含め、色々な方法で広報を行ってきたが、WWW で情報を提供するのには緊急度の高い使命であるという指摘があった。そのため、プロジェクト内の同様なニーズを含め、以下の目的をもって NIR Task Force を設立することになった。

- World Wide Web などを用いて、WIDE プロジェクトの活動を公開する
- WIDE プロジェクトでの研究に関連して、研究者間の情報交換のために、WWW サーバを構築、運用する
- WWW の日本語化、国際化について、問題点、実現可能性などを検討した上で、インプリメントを行う。

ここで、Working Group の形態ではなく、Task Force の形態をとったのは、必要性が差し迫ったものであったためである。

この報告書では、まず WWW システムについて説明し、その後、WIDE プロジェクトの WWW サーバについての説明を一通りする。関連事項として国際化についての説明を行なったあと、今後の展望について述べることにする。

---

<sup>1</sup>これは筆者の私見であるが、World Wide Web は、我々の想像を越えた速度でポピュラーになってしまったために、WIDE 内においても、対応が追い付いていない状況が多方面でみられていると思う。

## 第 2 章

# ネットワーク上での情報提供

### 2.1 WWW システムとは

WWW(World Wide Web) システムは、インターネット上で利用できるクライアント-サーバ型の情報サービスである。WWW システムのクライアントは、WWW ブラウザと呼ばれ、GUI を備えたものからテキストベースのものまで多種実装されている。

このシステムでは、情報 (画像や音声も含む) をページと呼ばれる小さな単位で管理しており、ページを相互に参照を行うことでハイパーテキスト文書を実現する。

各ページは、HTML(Hyper Text Markup Language) フォーマットで記述することで、静止画像を埋め込んだマルチメディア文書とすることができる。

また、HTML 文書内から参照されるのはページに限定されているわけではなく、動画像や、音声データ、その他のバイナリデータも参照できる。

### 2.2 URL(Uniform Resource Locater)

WWW システムでは、参照する情報コンテンツに一定の命名規則に従った名前をつけて管理されている。ここで用いられる命名規則が URL である。URL の構文は次のようになっている。

プロトコル名://ホスト名:ポート番号/パス

(例) `http://www.wide.ad.jp/index-j.html`

URL の先頭に記述するのは、通信プロトコル名で WWW システムの場合には、後述する HTTP(Hyper Text Transfer Protocol) を用いているので、対応する “http” を記述する。

次にアクセス先ホスト名を記述する。この例では WIDE project の WWW サーバである “www.wide.ad.jp” を指定している。

この例にはないが、ホスト名の後にポート番号を指定できる。指定がない場合には、デフォルトの “:80”(ポート 80 番を使用) が指定されたとみなす。

最後にサーバのルートディレクトリからのパスを指定する。上記の例では、ルートディレクトリにあるファイル “index-j.html” を参照している。

## 2.3 HTTP(HyperText Transfer Protocol)

普段ユーザの目に触れることがないが、WWW ブラウザと WWW システムのサーバは、HTTP を用いて通信を行っている。

HTTP のコネクションでは、リクエスト毎にクライアント側から接続し、レスポンスを受けると接続を解除する。

HTTP のリクエストには以下のようなコマンドが用意されている。

コマンド名	引数	作用
GET	パス名	指定されたパスに対応するデータを返す
POST	パス名	指定されたパスに対応する場所にデータを渡す

図 2.1: HTTP リクエスト

HTTP は、テキストベースのインターフェイスを持っているので、次のようにしてサーバの動作を確認できる。

```

% telnet www.wide.ad.jp 80
Trying 133.4.11.2...
Connected to endo.wide.ad.jp.
Escape character is '^]'.
GET /index.html HTTP/1.0                                <-- リクエストの投入
                                                         <-- 空行の入力
HTTP/1.0 200 Document follows                            <-- サーバのステータス
MIME-Version: 1.0                                       <-- MIME バージョン
Server: CERN/3.0                                         <-- サーバの種類
Date: Tuesday, 18-Apr-95 07:01:12 GMT                   <-- 現在時刻
Content-Type: text/html                                  <-- データの型名 (MIME)
Content-Length: 1387                                     <-- データのサイズ
Last-Modified: Sunday, 09-Apr-95 09:34:12 GMT           <-- 更新時刻

<TITLE>WIDE Project</TITLE>
<BODY>

(途中略)

</BODY>
Connection closed by foreign host.

```

## 2.4 HTML(HyperText Markup Language)

WWW システムで用いられるメニューに相当するページは、HTML フォーマットの文書で構成されている。

HTML は、構造化文書を実現するための記述言語規格 “SGML” のタグ記述方式を取り入れ、参照 (Anchor) タグや、埋め込み (Inline) イメージタグなどの拡張を行ってハイパーテキストを実現できるようにしたページ記述言語である。

タグ記述方式とは、“<タグ名>” のようなタグによりしるし (マーク<sup>1</sup>) をつけてゆくことで、文書に属性を与えたり、“<タグ名>” と、“</タグ名>” という開始、終了タグで囲むことで文書に構造を持たせるものである。

```
<html> <head>
<title>HTML Sample document</title>
</head>

<body>
<h1>HTML サンプル文書</h1>

<menu>
<li>
<a href="http://www.wide.ad.jp/">
WIDE project のホームページへジャンプ! </a>

<li>
<a href="http://www.wide.ad.jp/wide/announce/index-j.html">
WIDE project からのお知らせ</a>
</li>
</menu>

<hr>
<address>
WIDE WWW server &lt;www-admin@wide.ad.jp&gt;
</address>
<!-- hhmts start -->
Last modified: Fri Apr 21 13:17:19 1995
<!-- hhmts end -->
</body> </html>
```

HTML で導入された “参照 (Anchor) タグ” は、次のようにして使う。

この記述では、“WIDE project のホームページへジャンプ!” の部分が表示され、URL “http://www.wide.ad.jp/” へのリンクが張られる。

---

<sup>1</sup>マーク (印) を文書につけてゆくことから、マークアップランゲージと言われる。

## 2.5 CGI(Common Gateway Interface)

WWW システムでは、あらかじめ用意されたファイルのみを提供するわけではない。

URL に対応する情報コンテンツを参照された時点で動的に作成することもできる。この仕組みを WWW システムではアプリケーション・ゲートウェイと呼ぶが、その実体は CGI スクリプトというちいさなプログラムで実現されていることが多い。

CGI スクリプトは WWW のサーバプログラムから起動され、要求されたジョブをこなし、ジョブの結果をクライアントに直接返信する。

要求されたジョブの受け取り/結果の返信のインターフェイスが標準化されており、これを CGI と呼ぶ。

NCSA 版 WWW サーバのキットには、WWW で今日の運勢を教えてくれる次のようなプログラムが含まれている。

```
#!/bin/sh

FORTUNE=/usr/games/fortune

echo Content-type: text/plain
echo

if [ -x $FORTUNE ]; then
    $FORTUNE
else
    echo Cannot find fortune command on this system.
fi
```

この CGI スクリプトでは、システムのコマンド “/usr/games/fortune” があればこれを起動した時の標準出力を、コマンドがインストールされていなかった場合にはエラーメッセージを WWW クライアントに返す。



## 第 3 章

### WWW サーバの整備

NIR-TF の第一の目的は、WWW による WIDE プロジェクトの情報公開である。このため、WWW サーバを立ち上げ運用を行っている。運用には、www-admin という、NIR-TF 内のグループによって行われている。



図 3.1: WIDE Project ホームページ

www-admin は、各ワーキンググループの協力を得ながら、以下の情報を用意し、情報を公開している。

#### 3.1 WIDE プロジェクトの成果の公開

WIDE project における研究の成果は、毎年共同研究を行っている組織向けに作成される報告書、および蓄積された情報や、開発したソフトウェアの公開という形で公開されている。

これらの情報やソフトウェアは今までは NetNews や印刷媒体などでお知らせしてきたが、WWW によって公開することで、情報はもとより、関連するソフトウェアや図表も併せて公開出来るようになった。

## 3.2 WIDE プロジェクトに関連した情報の公開

この欄でも何度か説明があったと思うが、WIDE プロジェクトは、その研究活動の基盤として、WIDE Internet を構築、運営している。

残念ながら、WIDE プロジェクトの活動や、WIDE プロジェクトの目的や性格についての誤解があり、それによって、間違った報道がされたことさえあった。これの原因の一つとしては、WIDE プロジェクトが今何をやっているのか、WIDE プロジェクトの活動の中心であるインターネット上で提供、公開出来ていなかったのが上げられるであろう。

そのような点から、以下のような情報をまとめて公開している。

- WIDE プロジェクトに関する一般的な情報

- WIDE プロジェクトについて

WIDE プロジェクトについて説明している。

(<http://www.wide.ad.jp/wide/index.html>)



図 3.2: About WIDE

- WIDE Internet についての説明

Backbone, NOC, 接続サイト (および接続履歴) など。

(<http://www.wide.ad.jp/backbone/index.html>)

- ワーキンググループ (WG) タスクフォース (TF) の説明と、各 WG/TF それぞれについての説明 (設立趣意) や研究内容について。

(<http://www.wide.ad.jp/wg/index.html>)

### 3.3 WIDE プロジェクト 研究者間の情報交換

研究者間の情報交換の利用においては、以下のような事項があげられる。

- 研究会の発表資料の配布

これは事前に用意されるので、参加者は自分で印刷して研究会に参加する (WIDE 内公開)

- 半年に一度行われる合宿に関連した情報の集約化 (WIDE 内公開)
- 各ワーキンググループごとの情報集約

その他、必要に応じて、WIDE プロジェクトの活動に関連したサーバや、共同研究を行っているグループのサーバなどへのリンクなども作成している。

加えて、この後で述べるように、[www.wide.ad.jp](http://www.wide.ad.jp) の利用状況を自動集計して公開している (<http://www.wide.ad.jp/server-stat/>)。

### 3.4 データの収集

WWW での情報公開を行う場合、公開する情報そのものの、収集及び加工が必要である。

WIDE Project においては、論文を始めとする殆どの資料が計算機を用いて作成されており、機械的な変換による、それらの情報の加工は容易である。

しかしながら、多くの情報はそれを作成した者が個人的に管理しており、収集の手間から印刷された資料を再度入力することもあった。また、報告書の内容など WIDE 内部で電子的に公開されている情報も多いが、それらの情報のうちどの範囲までを公開して良いかどうかの判断も WWW 担当者だけでは行えず、結局明らかに公開しても良い情報のみを公開する事が多くなってしまった。

WWW 公開後は、研究成果の一般公開の方法の一つとして WWW をとらえ、WWW 用に公開データを作成し、提供する研究者も増えつつある。今後はそのようなデータ作成及び公開の支援を強化する予定である。

### 3.5 ftp との関係

WIDE では WWW 運用以前から、非常に普及している ftp を利用して自由に情報を取り出せる、匿名 ftp と呼ばれるサービスを行っている。WIDE の匿名 ftp ではフリーソフトウェア等の他、WIDE の研究成果なども公開している。

殆どの TCP/IP ソフトウェアに ftp を利用するプログラムが付属している他、電子メールを利用して匿名 ftp 上の情報を転送するサービスなども存在している。そのため、利用者の環境を選ばない情報公開手段として匿名 ftp は古くから利用されている。

しかしながら、ftp ファイルの転送に主眼を置いた機構である為、情報を全てファイル名で識別する必要があり、初心者には扱いにくいなどの問題もある。

WWW で利用されている URL では、情報取得方法として ftp を指定する事が可能である。そこでこの機能を利用し、HTTP サーバで公開されているページから ftp サーバへのリンクを設定し、ftp で公開されている情報は ftp で一元管理できるようにした。

これにより、従来の ftp でのアクセスに加え、WWW ブラウザによる容易なアクセスも可能になった。また、情報の実体は一つなので、情報更新時にも 2 種類のサービスで公開される情報の同期を取る必要が無くなった。

### 3.6 データ更新の自動化

http サーバを運用する上で、データを最新のものに保つことは重要な問題である。そこで、サーバで公開するデータのうち、自動化可能なものは自動化し、管理労力を軽減することを試みた。

データ群のうち自動化可能と思われるものには以下があげられる:

1. サーバのアクセス統計情報
2. サーバ内の文書の更新情報
3. WIDE メンバーの名前、所属、e-mail アドレス、ホームページの URL
4. WIDE ネットワークの統計情報

以下、これらのそれぞれについて説明する。

#### 1. サーバのアクセス統計情報

http サーバのアクセス状況は、httpd によってログファイルに出力される。しかし、ログファイルはアクセス元のホスト名、アクセス時刻、アクセスのあった文書名を記録しただけのもので、可読性が低く、また適切な統計処理がなされていないため理解しにくい(図 3.3 紙面の都合で 位置で折り返しているが、それぞれ一行である)。

```
www.wide.ad.jp - - [09/Apr/1995:17:29:37 +0900]
    "GET /icons/sq-r.gif HTTP/1.0" 404 -
www.wide.ad.jp - - [09/Apr/1995:17:29:38 +0900]
    "GET /icons/item1.gif HTTP/1.0" 404 -
www.wide.ad.jp - - [09/Apr/1995:17:29:38 +0900]
    "GET /icons/item1.gif HTTP/1.0" 404 -
```

図 3.3: httpd のログファイル例

このため、ログファイルからアクセス状況を時間帯ごとや日付ごと、ページごとに集計し、サーバのアクセス統計情報を HTML 文書として出力するスクリプト群が既に多数開発されている<sup>1</sup>。www.wide.ad.jp では中でも代表的な wwwstat<sup>2)</sup> を用い、統計情報を公開している<sup>3</sup>。統計情報の更新は 1 日 1 行われる。また、統計情報は 1 カ月ごとにファイルが分かれており、毎月 1 日にファイルの切替え処理が自動で行われる。

残念ながらこれまで数度事故により統計情報が失われ、現在残っているのは 1995 年 5 月からのデータのみとなっている。

## 2. サーバ内の文書の更新情報

http サーバ内で更新された情報の存在をユーザに通知することは重要であるが、そのメンテナンスを管理者が行うには多大な労力が必要である。このため、http サーバで公開しているディレクトリツリーに置かれているファイルのタイムスタンプを調査し、最終更新時刻が最も最近のファイルを一定個選択し、その内容からタイトルを抽出するスクリプトを開発した。現在、www.wide.ad.jp では 1 日 1 度このページを自動更新し公開している<sup>4</sup>。

## 3. WIDE メンバーの名前、所属、e-mail アドレス、ホームページの URL

www.wide.ad.jp では、WIDE 研究者と WIDE ボードメンバのリストを公開している。この際、WIDE のメーリングリストサーバである sh.wide.ad.jp に置かれた両メーリングリストのメンバリストと、http サーバ上のメンバリストの間の整合性を常に取りつづけることが必要である。この管理コストを抑えるため、sh.wide.ad.jp のメンバリストをもとに www.wide.ad.jp のメンバリストのページを自動生成するスクリプトを作成し運用している。データの更新手順は以下の通り:

- sh.wide.ad.jp 上の sendmail と SMTP プロトコルによる通信を行い、メーリングリストのメンバリストを得る。
- メンバリストに記録されている e-mail アドレスをキーに名前データベースファイルを検索し、名前の漢字表記、ローマ字表記、および個人ホームページの URL を得る。
- メンバリストに記録されている e-mail アドレスのドメイン名部分をキーに、JP-NIC のドメイン名リストから作成した組織名データベースファイルを検索し、所属組織名前の漢字表記、ローマ字表記、および組織のホームページの URL を得る。

<sup>1</sup>[http://www.yahoo.com/Computers/WorldWideWeb/HTTP/Servers/Log\\_Analysis\\_Tools/](http://www.yahoo.com/Computers/WorldWideWeb/HTTP/Servers/Log_Analysis_Tools/)

<sup>2</sup><http://www.ics.uci.edu/WebSoft/wwwstat/>

<sup>3</sup><http://www.wide.ad.jp/server-stat/>

<sup>4</sup><http://www.wide.ad.jp/recent-updates.html>

- 必要に応じて行を整形し出力する。個人ホームページおよび組織ホームページの URL がある場合には html によるタグづけを行う。

これらのページは以下からアクセス可能である:

- <http://www.wide.ad.jp/personal/board.html>
- <http://www.wide.ad.jp/personal/board-j.html>
- <http://www.wide.ad.jp/personal/list.html>
- <http://www.wide.ad.jp/personal/list-j.html>

#### 4. WIDE ネットワークの統計情報

[www.wide.ad.jp](http://www.wide.ad.jp) では、STAT working group の成果であるトラフィックの統計情報を公開している<sup>5</sup>。しかし、現在データの更新は全て手作業で行われている。今後 STAT working group と協調し、自動化を行うことが望まれる。

今後の課題としては、以下があげられる:

- より多くの情報の更新を自動化する

現在(および将来も)、[www.wide.ad.jp](http://www.wide.ad.jp) の管理を専任で行っているオペレータはおらず、管理は全て WIDE 研究者の一部のボランティア活動に頼っている<sup>6</sup>。管理コストを下げ、しかも情報の鮮度を失わないために、可能な限り情報収集/公開の自動化を図りたい。また、そのためのツール類が他サイトにとっても有用なら、広く世間に配布していきたい。

- httpd のアクセス制限との統合

httpd の設定によりアクセス制限を行っているページの URL は自動更新されたページに掲載しないことが望ましい。現在、このような制御は各スクリプト内で手動で行われているが、今後各スクリプトに httpd の設定ファイルを読み込む機能を付加し、アクセス制限に不整合が発生しないようにする必要がある。

---

<sup>5</sup><http://www.wide.ad.jp/wg/stat/nnstat/index.html>

<sup>6</sup>[www-admin@wide.ad.jp](mailto:www-admin@wide.ad.jp)

## 第 4 章

# WWW の日本語化

### 4.1 WWW の国際化についての研究

コンピュータのアプリケーションを用いる場合、利用される各国において、ソフトウェアがその国で自然に用いられている言語で使うことが出来ることが必須であるのは、論を待たない。

昨今では、Windows や Macintosh といった環境においても、システムのレベルで各国語化が自然に行われるようになった。これにより、パソコンやワークステーションのユーザにとっては、本来の言語を用いることを、ごくあたりまえに行えるようになってきている。

しかし、インターネットはその発展の過程から、英語圏での利用を前提に考えられている点が多い。そのうえ、国際的なやりとりが柔軟に行えて初めてメリットがあるシステムなのに国際化という点では、かなり遅れていると言えよう。一つの原因としては、インターネットが、まだまだ欧米圏主導である点が多いという点が大きいためであろう。これは、欧米圏の人々にとって、各国語化はそれほど重要な問題では無いために、差し迫った意識を持っていないためとも言える。

このような状況で、日本の人々にとって、さらに、非欧米言語圏の人々にとって自然に使う事が出来るような環境を作るためには、自分たちによって、かなり努力しなければならない。

NIR-TF では、このような状況から、WWW の国際化についても取り組んでいる。取り組みの一環として、NCSA Mosaic の日本語化、そして国際化をサンプルインプリメンテーションとして作成している。

サンプルインプリメンテーションは、まず XMosaic を改変することで始めている。コードの扱いについては、電総研の半田氏らによる多国語化された GNU Emacs、Mule と同じ方法をとっている。

将来的に、このインプリメンテーションを公開する予定ではあるが、時期的にはまだ未定であり、今回の報告にも具体的なインプリメンテーションを報告するには至らなかった。

これらの各国語化に関しての意識は、そのまま漢字コードと国際化コードに関連したト

ピックと強く関連している。ここでは Unicode についてとくに意見を述べるわけではないが、複数バイト文字コードを用いざるえない言語を用いるものとしては、1バイト文字コードを用いている人々の理解が足りないように思えてならない。

こういった事項についても、関連各方面に対する説明をするなどといった努力を通して、理解を深めてもらえるような努力をすることが、肝要である。今後の大きな課題の一つであろう。



## 4.2 NIA-SIG と、今後の活動

今回の報告は、現況を反映し、WWW システムにかたよったものになってしまった。しかし、WWW システムは恒久的に使えるほど洗練されたシステムではない。これは、データ転送プロトコルについても、それで利用できるアプリケーションの守備範囲からいってもそういえる。

たとえば、データ転送という点でみてみよう。現在は、データ転送のほとんどが静止画像などの「大きめなデータ」の転送が主なので問題が深刻化していないが、MPEG フォーマットの動画データなどの「巨大なデータ」の転送が頻繁に行われるようになったら、現在のプロトコルでは回線幅がいくらあっても足りなくなってしまう。

また、WWW システムでは転送プロトコルだけでなく、システムの構成自体にも問題がある。それは、WWW システムは情報コンテンツ管理コストの分散には成功しているが、情報生成のコストの分散ができていないことである。

効率的な情報収集システムや、情報の骨子だけを転送し、端末にてレンダリングを行なうようなシステムの導入で WWW 型のマルチメディア情報システムは、実用的なレベルになりえるだろう。

このような「理想的な情報システム」を構築するためには、マルチメディア情報のありかたについても考えなくてはならない。

これらの議論や、今年度の経験から、基本的なアプリケーションに対する WIDE 自体でアプローチを行なうことが重要であることが確認されたように思える。以下の事項が重要であろう。

- WWW 単体で出来ることにたいして、不満が多い。
- 研究をするためには、ソースコードがなければ問題を生じるが、これの入手に手間取るのは不幸である。ソースコードを得るために、とある機関とのやりとりを行なっていたが、最新の情報をタイムリーに得られないという点では非常に不満がある<sup>1</sup>。
- 他人によって作成されたプロトコルを不満を持ちながら用いるのは、研究とは言えない。
- なるべく多くの材料を手にした状態で作業をしたい。

このような観点から、WIDE 内で、WWW の先になにがあるのかを考えながら、プロトコルからはじまり、システムのアーキテクチャ、インプリメントを含めた形でのアプリケーションストラクチャを検討してゆくことになった。この議論は、新たに設立される NIA(Networked Information Architecture)-WG を設立し、検討することになった。

一方、NIR-TF は、WWW サーバをとりまく環境の整備に重点をおいて、今後も実作業を継続してゆく予定である。

<sup>1</sup>実際、ソースコードなしで Routing の研究をやるのに極めて近い

