

ネットワーク (4) WWW

かつらだ まさし
桂田 祐史

2002 年 5 月 16 日

1 連絡事項

- 先週の授業 (インターネット講習会) を受講した人は、情報科学センターのアカウントで、アクセス・レベル 2 の資格を得られているはずですが (例えば、学外の WWW ページが読めるなど、「アプリケーション・ゲートウェイ」経由でインターネット上のホストへのアクセスが可能になります)。確認は次のページで出来ます。

<http://www.meiji.ac.jp/mind/internet-license/>

2 WWW

(数年前までは、この授業で初めて WWW を使うという人もかなりいたのですが、今ではそういう人は珍しくなっていました。隔世の感があります。)

2.1 WWW とは

WWW (World Wide Web, 短く ^{ウェブ}Web と読まれることも多い) は、スイスの CERN (ヨーロッパ素粒子物理学研究所) で開発された情報発信 (情報検索、情報提供) システムであり、以下の特徴を持つ。

- (1) インターネット上のリソース¹を「リンク²」したハイパーテキスト³を記述できる HTML (hyper text markup language⁴) を扱う。

¹ コンピューターの世界では、「利用できるもの」を何でもリソース (resource, 資源) と呼ぶ。

² WWW のページ (ハイパーテキスト) の中で、別のインターネット・リソースを後述の URL で指定しておいて、マウスの操作一発でそのリソースを参照することを可能にした仕掛けのこと。ハイパーリンクとも言う。

³ ハイパーテキストという語そのものは、それ以前から使われていて、(1) 様々なデータを「参照」によって結びつけた、(2) (文字情報だけではない) マルチメディア情報を取り込んだ、という二つの条件を満たした、従来からのテキストを越えた新しいテキスト、と言った意味で使われていることが多い (hyper = ~を越えた、非常な)。Macintosh の HyperCard, TRON のダイナミック・ドキュメント、 \TeX の Hyper \TeX 拡張など。

⁴ HTML の例が見たければ、IE, Netscape などのブラウザの「表示」メニューから「(ページの) ソース」を選択すれば良い。

- (2) リソースの場所を示すために URL (後述) を用いる。
- (3) 主たる通信プロトコルとして HTTP (hyper text transfer protocol) を用いる (このプロトコル⁵自体は大変単純である)。
- (4) リソースとしてはテキストだけでなく画像、映像、音声などのいわゆるマルチメディア・データ⁶が扱える。

URL とは

URL は “Uniform Resource Locator” の略。直訳すると「統一的なリソース指示法」。インターネット上の各種リソースの「場所」を統一的な仕方で指定するフォーマット、あるいはそのフォーマットによる実際の「場所の指定」のこと。

<code>http://www.meiji.ac.jp/index.html</code>	明治大学のホームページ
<code>http://www.sony.co.jp/index.html</code>	ソニーのホームページ
<code>http://www.nasa.gov/index.html</code>	NASA のホームページ
<code>http://www.asahi.com/index.html</code>	朝日新聞のホームページ
<code>http://home.netscape.com/</code>	Netscape Comm. のホームページ
<code>ftp://ftp.iij.ad.jp/</code>	有名なプロバイダー IIJ の ftp サーバー
<code>gopher://gopher.ncc.go.jp/</code>	国立がんセンターの gopher サーバー
<code>file:/home/mk/internet.tex</code>	ローカル・ホストの “/home/mk/internet.tex” というパス名のファイル

などの例からも大体の「形式」は分かるが、一般には

スキーム名:ホスト名.ドメイン名[:ポート]/パス名[?検索条件]

という形式になっている⁷。スキームは、データの転送の仕方を表す (“http” はハイパーテキスト転送プロトコル、“ftp://” はいわゆる ftp (ファイル転送プロトコル)、“file” はローカルなファイルへのアクセス)。コンピューター (ホスト) を指定するには、ドメイン・ネーム・システム (DNS, 付録で解説) が用いられる。パス名はそのコンピューターの中で、ファイルの場所を示す「名前」のことである (この言葉そのものは UNIX に由来している)。

(最近では URL を拡張した URI (“Uniform Resource Identifiers”) という言葉が使われることも多くなった。現在では、URL は、URI のうち、主たるアクセス手段を識別の手法とするものを指す非公式な概念である、と説明するものだそうです (<http://www.kanzaki.com/docs/html/htminfo-uri.html> を見よ)。

2.2 WWW ブラウザー

WWW ページを読む (ブラウズ) するためのソフトウェアを WWW ブラウザーと呼ぶ。

⁵通信の世界では、通信のための取り決め、と言った意味。http = Hyper Text Transfer Protocol, ftp = File Transfer Protocol, etc.

⁶現在では、画像、映像、音声が使えるのは当たり前になってしまっていて、かえって「マルチメディア」は死語になったように思えるのは気のせいでしょうか。

⁷[] は省略可能であることを表す。

http (WWW) 以外の ftp, WAIS, Gopher, E-mail, Network News など、様々な機能を備えた統合的なソフトウェアになっているものが多い。

大抵は GUI (graphical user interface) を持ち、マウスの操作で色々なことが出来るが、CUI のテキスト・ブラウザもある (文字情報が主たる内容のページが案外多いので、かなり実用的である。特に広告がゴタゴタしているページではむしろ快適とも言える...)

Internet Explorer 以来、企業戦略的に (一見) 無料で配布されているものも多い。

WWW ブラウザーの例

Mosaic	NCSA 開発 (1993) の最初の定番 WWW ブラウザー。 (Netscape, IE の間接・直接の基礎となった。)
もじら Netscape	Netscape Communications Corp. 製の、ある時期の定番ブラウザ。 様々な OS に移植されている。
Internet Explorer	Microsoft 製。現在シェア一番。
Lynx	テキスト型ブラウザ
w3m	日本製テキスト型ブラウザ
Mozilla	Netscape の親戚であるオープン・ソースのブラウザ。
Opera	最近シェアを伸ばしつつある。

情報処理教室のパソコンでは

情報処理教室の Windows 環境には Internet Explorer, Netscape がインストールされている (いずれもスタートメニューから起動可能。終了するには“ファイル”メニューから“終了 (x)”を選択すればよい。)⁸。

情報処理教室のパソコンはインターネットに直接接続はされていない。インターネット講習会を受講した人に限り「MIND アクセスレベル2」の利用が可能で、アプリケーション・ゲートウェイ経由でインターネット上の WWW ページにアクセス出来るようになっている。

事前に [スタート] メニューの [レベル2 利用資格認証] を選択・実行し、ユーザー名、パスワードを入力する必要がある。

2.3 WWW ページのアクセスの仕方

見たいページの URL を指定してやればよい。そのための方法としては、以下の三つの手段がある。

1. URL を指定してのページのアクセス URL を入力してページを指定するには、次の 1, 2 のいずれかを行なう。
 1. “アドレス (D) あるいは“場所: ” の右にある枠に (マウスでクリックして選択して) キーボードから入力して、リターン。

⁸今年度の情報科学センターのシステムの UNIX (Solaris) 環境では、WWW ブラウザーはサポートされていない。

2. 次のいずれかで現れるウィンドウに、キーボードから URL を入力して、適当なボタンをクリックする。
 - (a) [ファイル]メニューから [(ページを)開く(O)]を選択
 - (b) キーボードで Ctrl-O (Ctrl キーを押しながら別のキーをタイプすることにより、Ctrl-何とか、という特殊な文字が送信できる。)を入力

なお、URL の先頭の `http://` や末尾の `index.html` は省略可能である。

2. リンクをたどる 既に表示してあるページの中のボタンをクリックして、リンクをたどる。
3. ブックマーク (お気に入り) から選択 これについては次小節で説明する。

2.4 ブックマークの使い方

しばしば訪れたいページは記録しておいて (Netscape では「ブックマーク」に追加、IE では「お気に入り」に追加)、一覧からマウスで選択することが出来る。

いくつかのページを尋ねて、(ブックマーク/お気に入り)に登録してみよう。Netscape では、[ブックマーク]メニューの先頭にある[ブックマークを追加]を選択する。IE では[お気に入り]の[追加]ボタンをクリックする。

Netscape のブックマークの編集法

[ブックマーク]メニューで[ブックマークを編集]をクリックすると、Bookmarks ウィンドウが現れる。ここに[ファイル]、[編集]、[表示]というメニューがある。

フォルダによる階層的整理 フォルダ (Folder, 書類はさみ) を作るには、[ファイル]メニューから [新しいフォルダ(F)]を選択し、現れたウィンドウに対し名前などを入力。

ブックマークの削除 まず削除したいブックマークをクリックして選択しておいてから、[編集]メニューの[削除]をクリック。

ブックマークの移動 (順序入れ換え) ブックマークをマウスでドラッグすれば良い。

ブックマークの名前 (見出し) の変更 まず変更したいブックマークをクリックして選択しておいてから、[編集]メニューの[ブックマークのプロパティ]をクリックして現れるウィンドウの「名前」フィールドを変更し、[確認]ボタンをクリックする。

2.5 検索エンジン

無数にある WWW ページの中から、自分が必要とする or 興味を持っている情報を探すにはどうしたら良いだろうか? 人から面白いページの URL を教わるというのももちろん結構だが、実はそのものずばりホームページ検索用のホームページがある。そこで働いている検索用プログラムを検索エンジンと呼ぶ。

検索エンジンは、元々は

1. 人間が人力で収集・整理したもの (ディレクトリ系)

例えば ^{ヤフー}Yahoo (<http://www.yahoo.co.jp/>) が有名。

2. プログラムが力任せに探索・機械的整理したもの (ロボット系)

例えば Google (<http://www.google.com/>), Lycos (<http://www.lycos.co.jp/>), goo (<http://www.goo.ne.jp/>) など。

の二つに分類されたが、現在は、かつてはディレクトリ系の代表とされた Yahoo 等でもロボット系を補助として利用するようになってきた。

検索エンジンの実際の使い方は、「百聞は一見にしかず」なので、ここでは省略する。

検索エンジンは栄枯盛衰が激しいので、評判に注意。「日本の Search Engine のリスト」という WWW ページ (<http://www.ingrid.org/w3conf-bof/search.html>) が参考になる。

2.6 その他

(もう説明するまでもない?)

- 一つ前のページに戻るには [戻る] ボタンをクリック。何回か戻れるが [次] あるいは [進む] で逆方向に動ける。
- [ジャンプ] あるいは [履歴] メニューには、読んだページの一覧表があるが、ここから特定のページを選んでジャンプできる。
- ネットワークが混雑しているときなど、しばしば転送が途中で終了してしまうことがある。こういう場合は“再読み込み” あるいは“更新” と書いてあるボタンをクリックすると再度読みに行く。(“再読み込み” をクリックしても、実際にはキャッシュと呼ばれる一時記憶を読むだけで、その WWW ページにアクセスしないこともある。こういう場合は、Shift キーを押しながら“再読み込み” をクリックする。)
- 現在表示しているページの中で検索を行いたいときは“編集” メニューの“ページ内を検索 (F)” を選択する、Ctrl-F を押す。
- 見ているページを印刷するには、[印刷] メニューを呼び出せば簡単。なお、後から印刷出来るようにファイルに保存しておくことが可能である。
- 見ているページをファイルに保存するには [ファイル] メニューから [名前を付けて保存 (S)] を選択する。保存の形式として、ソース、テキスト、PostScript の 3 種類が選べる。(リンクしてあるものを保存するには Shift キーを押しながら左ボタンでクリックするのが早い。)
- 画像をファイルに保存するには、画像のところにマウス・カーソルを持って行って、右ボタンをクリックし、[画像を名を付けて保存] を選択する。保存した画像ファイルは、画像ビューアーやブラウザ自身で表示できる。
- 一人のユーザーが同時に複数の Netscape/IE を起動してはいけない。複数のウィンドウが必要な場合は、リンクのところでも右ボタンを押し、現れたメニューから [新しいウィンドウで開く] を選択する。
- 日本語の文章が入っていると思われるページで、「文字化け」が起きたら⁹、まず“再読み込み”を試す。それでも駄目だったら [表示] メニューの [文字コードセット] あるいは [エンコード]

⁹日本語の文字を表わす文字コードには、色々なものがあるが、WWW ページでは標準化がされていない(望ましい文字コードというものはあるが、守っていないところが多い)。「行儀の良い」WWW ページでは、そのページで用いている文字コードをきちんと指定してあるが、そうでない WWW ページは少なくない。そういうページを読むとき、WWW ブラウザーは、そのページで使われている文字コードの種類を自動判別しようと試みるが、原理的に完全に出来るものではないため、時々失敗する。そのとき「わけのわからない表示」— 文字化け — になる。

で日本語 encoding を色々試してみる (本来 AutoDetect (自動検出、自動選択) で大丈夫なはずだが)。

- ftp を利用するには、URL として “ftp://ホスト名.ドメイン名/パス名” を指定する。後はマウスの直感的な操作で済む。

例: “ftp://ftp-asjp.netscape.com/pub/communicator/japanese/”

2.7 注意事項

- ページの読み込みが遅いのは、そのホストまでのネットワークが混んでいる可能性が高い。はなはだしい場合は以下のいずれかの方法で対処しよう。
 - (1) そのページのアクセスをあきらめる
 - (2) 代替ページ (mirror site) がある場合はそちらをアクセスする
 - (3) 後で空いている時間帯を見計らってアクセスする
- WWW を使ったアンケート、掲示板の取り扱いには特に注意する。
 - 自分の個人情報はむやみに書き込まない。
 - 他人を誹謗中傷する内容など書き込まない。

3 レポート課題3

次の課題の内容は次のページに書いてある。自分で読んでレポートを提出せよ。締め切りは5月29日とする。

<http://www.math.meiji.ac.jp/~mk/syori2-2005/report3.html>

A おまけ WWW 用語集

大事なものは本文で説明したので、ここにあるのは、残り物です。

ホームページ WWW では、個々のハイパーテキスト・データを WWW ページ (あるいは単にページ) と呼ぶが、ある一つの WWW サービスの玄関に相当するページをホームページと言う (「トップページ」という呼び方もある)。

(明治大学のホームページ、明治大学数学科のホームページ、桂田祐史のホームページ、などなど。)

最近では玄関に相当しないページもホームページと呼ばれることが多くなった (WWW そのものをホームページという人まで現れた。言葉の濫用?)。

ネット・サーフィン WWW のページをリンクを辿ってあちこち渡り歩くことを、気取ってこう言う。もともとは、ケーブル・テレビが普及したアメリカの「チャンネル・サーフィン」という言葉の連想から来た言葉だとか。

B ドメイン・ネーム・システムの常識

要点: インターネット上のコンピューターを指定するには DNS という仕組みを使う

手短かに説明すると、

DNS (domain name system, あるいは domain name service の略) とは、ホスト名、ネットワーク名の分散型データベース・システム。ホストをドメインと呼ばれるグループに分ける階層を作成し、インターネット上の各コンピューターに対して、ドメイン名を割り当てる仕組み。(インターネット上のホストとネットワークの数は膨大なものになり、一箇所で集中管理することは難しいので、それを解決するために作られた。) DNS のサーバーをネーム・サーバーと呼ぶ。

以下はゆっくり解説。

TCP/IP においては、IP アドレスと言う 4 バイトの数値で、コンピューターを識別するが、これは人間が使うのには不便なので、

- 各コンピューターにホスト名 (host name) と呼ばれる名前をつける
- ホスト名と IP アドレスの対応表を用意する

ことで、通常は、人間はホスト名を使い、コンピューターは IP アドレスを使い、必要に応じて両者を相互に変換する、という仕組みになっている (例えば、数学科で一番大きなコンピューターは oyabun というホスト名を持ち、その IP アドレスは 133.26.132.30 である)。

MIND (明治大学内ネットワーク) に接続する各コンピューターは、ホスト名のみによる識別が可能になっている。(例えば “waltz” というホスト名を持つコンピューターは一台しかない。MIND のすべてのコンピューターはホスト名 “waltz” のコンピューターがどこにあるか=waltz の IP アドレスがいくつであるか、知っている。)

ところが、インターネットには、無数のコンピューターがあるので、簡単なホスト名だけで、すべてのコンピューターを識別することは困難である。そのために作られたのが、domain name system (DNS) という仕組みである。

DNS では、

1. インターネットを複数のドメインと呼ばれるものに分割して、ホスト名と IP アドレスの対応表はドメインごとに管理する分散データベースとして実現する。
2. インターネット上のコンピューターは、ホスト名にドメイン名を添えることで一意的な名前をつけることにする。

のような仕組みになっている。ちなみに明治大学内のネットワークには “meiji.ac.jp” というドメイン名が付けられている。

実際には、meiji.ac.jp というドメインはいくつかのサブドメインに分割して管理されている (つまり DNS の名前の空間は階層構造を持っている)。サブドメイン名としては、isc (情報科学センター), cs (情報科学科), math (数学科), mind (MIND) などがある。例えば、情報科学センターにある sagami2 は sagami2.isc.meiji.ac.jp という名前。

C 学内 URL メモ

明治大学ホームページ 一度は目を通しておくべき明治大学の公式ホームページ。

<http://www.meiji.ac.jp/>

明治大学情報科学センターホームページ 情報科学センターからのお知らせや、利用のための便利情報。

<http://www.isc.meiji.ac.jp/>

明治大学数学科ホームページ 一度は読んでみよう。大学院生や阿原先生が頑張って作っている。
<http://www.math.meiji.ac.jp/>

数学科情報処理 II のページ 「以前読んだプリントに書いてあったあれを読みたい」という場合に。
<http://www.math.meiji.ac.jp/\textasciitildemk/syori2-2005/>

明治大学総合情報ネットワーク MIND のホームページ MIND の仕組み、掟を知りたい時に。
<http://www.meiji.ac.jp/mind/>

インターネット講習会の資料 「インターネット講習会」の復習をしたい時に。
<http://www.meiji.ac.jp/mind/seminar/shikaku2002/>

MIND モバイル・アカウント・オンライン手続きシステム モデムや ISDN ターミナル・アダプターを使って、自宅のパソコンを明治大学のネットワークに PPP 接続したり、自分のノートパソコンをモバイル情報コンセントに接続するための資格を得る手続き。
<http://mobile.mind.meiji.ac.jp/>

ホームページの立ち上げ方・超・簡易手引き 情報科学センターのユーザーが自分のホームページを作るための方法を説明してある。<http://www.isc.meiji.ac.jp/faq/homepage.html>

D おまけ: telnet で WWW サーバーとお話ししよう

```
telnet WWWサーバー名 80
GET /パス名
```