情報処理 II 第 3 回

ネットワーク (3) WWW

桂田 祐史 (かつらだ まさし)

2005 年 5 月 12 日

ホームページは http://www.math.meiji.ac.jp/~mk/syori2-2005/

(以前は使い方や「最新情報」を伝えていたのが、いつの間にか「歴史」の説明が増えてきている...)

1 連絡事項

- アンケートの結果、これまでインターネット講習会を受けた / 受けないは半々 (昨年は 34/13, 一昨年は 24/23, 一昨年並みでやや後退？) であった。次回 (5 月 19 日) はインターネット講習会をする。
  - 既にインターネット講習会を受けた人は出席する必要はない (出席しても構わない)。
  - 10 分以上の遅刻は認めない[1]、早退も認めない (出席したと認めない)。

2 WWW

かつては[2]、この授業で初めて WWW を使うという人も多かったが、今ではそういう人は珍しくなってしまった。隔世の感がある。また、この情報処理 II 以前にインターネット講習会を受講している人も多くなった。今回 WWW の説明をするわけだが、「ソフトの使い方」についての説明は省略する。自分で調べるなり、友人に尋ねるなりして対処すること。

2.1 要点

- 思い切って単純化すると、WWW 技術そのものは次の 4 つからなる。
  1. HTML (hyper text markup language)
  2. HTTP (hyper text transfer protocol)

[1] 講習会に 70 分強かかるので 10:40 から講習会を始めることと、講習会自体への遅刻は一切認められないため。

[2] この講義で初めて WWW の解説をしたのは 1996 年だった。

3. URL (uniform resource locator)
4. DNS (domain name system)
   付録 B に簡単な説明がある。

- 個人的には WWW にまつわる二つの奇跡があると感じている。
  1. WWW は案外単純な仕掛けだが、これだけうまく動いて有用であるという事実
  2. Google に代表されるロボット系検索エンジンの成功

2.2 WWW とは

WWW (World Wide Web, 短く Web(ウェッブ) とも読まれる) は、スイスの **CERN** (ヨーロッパ素粒子物理学研究所[3]) で Tim Berners-Lee 氏[4]が論文閲覧システムとして開発したもの (1990) が原点となっている情報発信 (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[5]を「リンク[6]」したハイパーテキスト[7]を記述できる **HTML** (hyper text markup language) をデータの記述言語として用いる。

(HTML が実際どういうものかについては、次節を参照せよ。)

(2) 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[8]自体は大変単純である。単に指定された (自分が持っている) データを送るだけと言える。

[3]2002 年にノーベル物理学賞を受賞した小柴昌俊氏なども活躍した素粒子物理研究のメッカ。

[4]HTML, HTTP, URL を創った。「ティム・バーナーズリー博士」(`http://www.ibarakiken.gr.jp/www/world/`) が面白い。

[5]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。

[6]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。

[7]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 HyperTEX 拡張などがハイパー・テキストの例としてあげられる。

[8]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.

直接 WWW サーバーと話してみよう

UNIX 環境から telnet コマンドで WWW サーバー (例えば数学科の www.math.meiji.ac.jp) の 80 番ポートにアクセスしてみよう)。

```
a308-01% telnet www.math.meiji.ac.jp 80
Trying 133.26.140.28...
Connected to mathweb.mind.meiji.ac.jp.
Escape character is '^]'.
GET /index.html           ← これが唯一のコマンド入力 (大文字に注意！)
<HTML>
<HEAD>
  <META NAME="GENERATOR" CONTENT="Adobe PageMill 3.0J">
  <META HTTP-EQUIV="Content-Type" CONTENT="text/html;CHARSET=x-sjis">
  <TITLE>Department of Mathematics, Meiji University </TITLE>
</HEAD>
<BODY BACKGROUND="bg0.gif" TEXT="#660066">
... (省略) ...
</table>
</body>
</html>
a308-01%
```

つまり単に GET ファイルのパス名 として指定されたファイル (上の例では index.html) を送ってくるだけである。ブラウザーは送られたデータを適切に解釈して必要な処理をすることになる。
(注意: 原理的には Windows 環境の DOS 窓でも試せるはずだが、日本語が表示できない可能性が高いので奨められない。)

(3) リソースの場所を示すために **URL** (すぐ後に説明) を用いる。

(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・データが扱える[9]。

URL とは

URL は "Uniform Resource Locator" の略。直訳すると「統一的なリソースの場所指定法」。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。

[9] テキスト以外のものが特別だった時代は昔話となったので、もう「マルチメディア」も死語かもしれない？

URL の例

`http://www.meiji.ac.jp/index.html`	明治大学のホームページ
`http://www.sony.co.jp/index.html`	ソニーのホームページ
`http://www.nasa.gov/index.html`	NASA (アメリカ航空宇宙局) のホームページ
`http://www.asahi.com/index.html`	朝日新聞のホームページ
`http://home.netscape.com/`	Netscape Comm. のホームページ
`ftp://ftp.iij.ad.jp/`	有名なプロバイダー IIJ の ftp サーバー
`gopher://gopher.ncc.go.jp/`	国立がんセンターの gopher サーバー ついになくなった！
`file:/home/mk/internet.tex`	ローカル・ホストの “`/home/mk/internet.tex`” というパス名のファイル Windows なら `file:///z:/internet.tex` のような感じ。

これらの例からも大体の「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[10]。

- 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ある (`http://www.kanzaki.com/docs/html/htminfo-uri.html` を見よ)。

2.3　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。

http (WWW) 以外の ftp, WAIS, Gopher[11], 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。

---

[10] [　] は省略可能であることを表す。

[11] Gopher は 1980 年代にミネソタ大学で開発された分散型掲示板システムである。“Gopher” は北米地域に生息するネズミの一種 (ホリネズミ) で、辞書によるとミネソタ州人という意味もあり、ミネソタ大学のマスコットにもなっているそうである (ちなみに Disney の「くまのプーさん」にも登場する)。go for ... とかけて名付けられているとか。

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 (character user interface) のテキスト・ブラウザーもある (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[12])。

WorldWideWeb	世界最初の WWW ブラウザー (1990)。ティム・バーナーズリー氏が作成。
Mosaic	NCSA 開発 (1993) の最初の定番 WWW ブラウザー。
	(Netscape, Internet Explorer の間接・直接の基礎となった。)
Netscape	Netscape Communications Corp. 製の、ある時期の定番ブラウザー。
	様々な OS に移植されている。
Internet Explorer	Microsoft 製 (最初は Mosaic を改良した)。現在シェア一番。
Lynx	テキスト型ブラウザー。
w3m	日本製テキスト型ブラウザー。
Mozilla	Netscape から派生したオープン・ソースのブラウザー。
	最新版は Firefox と呼ばれ、人気上昇中。
Opera	最近シェアを伸しつつある。

(そろそろ携帯電話の WWW ブラウザーの話でも書くのだろうか...)

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、Linux 環境には Mozilla がインストールされている。

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。

Windows XP 環境では、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。

操作法は... もう説明無いと思われるので省略。

2.4 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 を直接入力する[13]。
2. 既に表示してあるページの中のリンクをたどる。
3. ブックマーク (お気に入り) から選択する。

[12]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...

[13]ファイル・メニューから「開く」を選択するか、キーボードから、IE なら Ctrl-O, Netscape なら Ctrl-Shift-L (何だこれは) をタイプする。あるいは「アドレス」をクリックして選択して、キーボードから URL を打ち込む。

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か？友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 WWW ページ検索用の WWW 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 Yahoo (ヤフー) (`http://www.yahoo.co.jp/`) が草分け。
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/`), Infoseek (`http://www.infoseek.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ィ系の代表とされた 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ている。

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。

いくつかある検索エンジンの評判については、「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... と言ってきたが、このところ Google の一人天下で、もはや意味のある評判は見られない。

Google では、検索キーワードを含むページの順位づけに工夫 (PageRank という特許になった技術) をしているが、大規模な数値計算で固有値問題を解いている (参考: おおざっぱには `http://www.math.meiji.ac.jp/~mk/shiminkouen/ohp/mgp00017.html`, 詳しくは馬場肇氏の『Google の秘密 — PageRank 徹底解説』[14] または同氏が数学セミナー 2004 年 6 月号に書いた「情報検索と行列」)。

2.6　その他

1. 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” をすると良い。
2. 見ているページをファイルに保存することもできる (「ファイル」メニューを参照)。
3. 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。

[14] `http://www.kusastro.kyoto-u.ac.jp/~baba/wais/pagerank.html`

3　試しに WWW ページを作ってみよう

明治大学の情報科学センター生田分室にアカウントを持つ人は、以下のように簡単にホームページを開設できる。

百聞は一見にしかずなので、試しにやってみることを勧める。

WWW ページを作るためのソフトが普及しているが[15]、原理を理解するためには、一度は自分の手で HTML ファイルを書いてみるのが良い。

なお、手っ取り早く HTML で書かれたファイルの例が見たければ、Internet Explorer, Netscape などのブラウザーの「表示」メニューから「(ページの) ソース」を選択すれば良い。

3.1　参考資料

もちろん書店でたくさんの参考書を見つけることができるが、とりあえず試すだけならば、以下のページに書いてある情報だけで十分であろう。

`http://www.meiji.ac.jp/mind/seminar/shikaku/330-i7.html`
『インターネット講習会資料 自分のホームページの開設方法』

`http://www.isc.meiji.ac.jp/faq/homepage.html`
『ホームページの立ち上げ方・超・簡易手引き』
(注: 情報処理教室で利用しているシステムでは、ホームディレクトリィは、Windows 環境では `z:` ドライブになる。)

3.2　ホームページ開設手順

`~/public_html` というディレクトリィを作って、アクセス・モードを他人からも検索可能なようにしておく。

初めに一度おまじない

```
a308-01% cd
a308-01% mkdir public_html
a308-01% chmod o+x . public_html
```

(Windows 環境にログオンしている場合は、TeraTerm Pro で samba00 等にリモート・ログインするのが簡単かもしれない。)

既存の HTML ファイルを参考に自分で一つ書いてみる。上記『ホームページの立ち上げ方・超・簡易手引き』にあるサンプル `http://www.math.meiji.ac.jp/~mk/syori2-2005/hogehoge-jis.html` (UNIX 向け), `http://www.math.meiji.ac.jp/~mk/syori2-2005/ms/hogehoge-ms.html` (Windows 向け — 修正しました) を用意してある。

[15] ちなみに、この授業の資料は LaTeX 2$_\varepsilon$ で書き、latex2html というプログラムで、html 形式に変換して作成している。

例えば emacs で書く

```
a308-01% cd public_html
a308-01% emacs hogehoge-euc.html &        名前は自由（拡張子は .html）
```

例えば秀丸で書く

z:¥public_html の下に書くことだけ注意すればよい。2005 年度のシステムでは z: ドライブが呼び出しづらいが、「ファイル」メニューの「名前を付けて保存」を選択して現われるウィンドウの「ファイル名 (N)」欄に [z:] [Enter↵] を入力すればよい。それから public_html フォルダを選び、「ファイル名 (N)」には例えば [hogehoge-euc.html] を入力して、[保存] ボタンを押す。

HTML については、

```
<html>
<head>
<title>This is a title.</title>
</head>

<body>
<h1>見出し</h1>

例えば <a href="http://www.asahi.com/">ここは朝日新聞</a> のように
してリンクを行なう。

<hr>
<address>
  <a href="mailto:mk@math.meiji.ac.jp">メイルはこちら</a>
</address>

</body> </html>
```

本当は文字コードの説明などをする必要があるが、それは次回以降に解説予定。

4 レポート課題 2

締め切りは現在 5 月 25 日の予定 (もしかしたらもう一週延ばします)。提出方法はこれまでと同じで、syori2@math.meiji.ac.jp 宛に送る。Subject (件名) は「情報処理課題 2」、自分が誰であるかきちんと書くなどの注意 (もう一度繰り返すと、本文の先頭に学年・組・番号・氏名を書く) も忘れずに。

次の質問に答えよ。

(1) ドメイン名を答えよ。
(a) 明治大学 (b) 株式会社 SONY (c) 明治大学理工学部数学科 (サブドメイン)

(2) 自分の情報科学センターのアカウントの E-mail アドレスは？

(3) MIND アクセス・レベル 3 のホストを一つあげなさい。

(4) 以下の言葉を使って WWW の仕組みを説明しなさい (3 ~ 4 行程度でよい)。
(a) HTTP (b) HTML (c) URL (d) DNS

(5) 普段検索エンジンを使っているならば、どの検索エンジンを使っているか、それを選んだ理由は何か、主にどういう用途に使っているか、教えて下さい。検索エンジンを使ってない場合はその理由を教えて下さい。

なお、研究課題として、Google の PageRank 技術についてのレポートを提出してくれても良い[16](締め切りは最後の授業まで)。

[16]面白いし、数学の分かりやすい応用の一つなので、調べてみるのは意義深いと思われる。

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で「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。

B.1　ホスト名と IP アドレス

TCP/IP においては、**IP** アドレスと言う 4 オクテット (=32 ビット) 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 “samba03” 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 “samba03” のコンピューターがどこにあるか=samba03 の IP アドレスがいくつであるか、知っている。)

B.2　ドメイン名

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。

nslookup コマンドで DNS に問い合わせ

```
oyabun% nslookup
Default Server:  ikuta-ns.mind.meiji.ac.jp
Address:  133.26.136.30

> www.meiji.ac.jp
Server:  ikuta-ns.mind.meiji.ac.jp
Address:  133.26.136.30

Name:    www.meiji.ac.jp
Address:  133.26.136.152

> www.sony.co.jp
Server:  ikuta-ns.mind.meiji.ac.jp
Address:  133.26.136.30

Non-authoritative answer:
Name:    vip.alt.ihp.sony.co.jp
Address:  210.139.255.221
Aliases:  www.sony.co.jp, www-sony.ihp.sony.co.jp

> 133.26.132.30
Server:  ikuta-ns.mind.meiji.ac.jp
Address:  133.26.136.30

Name:    oyabun.mind.meiji.ac.jp
Address:  133.26.132.30

> Control-D
oyabun%
```