最終更新日　2015.10.29

# html言語とWebの仕組み

　インターネットの中で最も重要な技術であるWebのしくみを理解します。

## Webブラウザ表示のしくみ

　インターネットエクスプローラ（以下IEと表現します）やFirefoxのアドレスに、例えば

http://www.nara-edu.ac.jp/guide/list/index.html

と入れると、www.nara-edu.ac.jpというWebサーバと通信を行い、そのサーバが保持する /guide/list/index.htmlというファイルを取ってきます。次に、そのファイルの内容を解釈してブラウザの表示画面の中に表示します。ファイルの形式はhtmlと呼ばれる形式になっています。テキスト形式なのでWordやメモ帳などで読むことも出来ます。IEの場合、「表示」→「ソース」と操作するとhtml形式のファイルの内容を閲覧することが出来ます。Firefoxの場合は「右クリック」→「ページのソースを表示」と操作して下さい。

（注）多くのサイトではファイル名を省略したとき、index.htmlがデフォルトのファイル名としてURLの末尾に付加されます。

## htmlとは

　最も単純なhtmlファイルを次に示します。

<html>

<head>

<title>ここにタイトルを書く</title>

</head>

<body>

ここに何か書く

</body>

</html>

　メモ帳などのエディタで上記のファイルを作成し、Webブラウザで読み込んで下さい（「ファイル」→「開く」あるいはファイルをWebブラウザにドラッグ＆ドロップする）。

　この例のように、htmlファイルは全体が <html> </html> で囲まれています。その中に <head> </head> で囲まれたヘッダと呼ばれる部分と <body> </body> で囲まれたボディと呼ばれる部分があります。ヘッダの中に <title> </title> で囲まれた部分があり、ブラウザのタイトルバーに表示される文字列として使われます。また、「お気に入り（bookmark）」に取り込んだときの見出しにもなります。bodyの部分がブラウザの表示領域に表示されます。bodyの部分を次のようにしてみましょう。

<body>

<h1>これはレベル1の見出しです。</h1>

<hr>

<p>

一つ目の段落。

</p>

<p>

二つ目の段落

</p>

</body>

　エディタでファイルを書き換えた後、フロッピーディスクのアイコンを押してファイルをセーブします。次に、ブラウザの「最新の情報に更新」あるいは「このページを再読み込みします」というバルーンヘルプが出るアイコンをクリックして下さい。更新してセーブしたファイルを読み込んで再描画します。この実験から分かるように、Webブラウザはhtml形式で書かれたファイルを解釈して表示する働きを持っています。

　<h1>　</h1>　をタグと呼びます。h1はレベル1の見出しを作る命令です。<h1> は開始タグでその後に内容を書き、</h1> が終了タグです。<p> は段落の開始を示す命令で、</p> で段落の終了を表します。次の段落との間で改行を行い、縦方向のスペースを挿入します。<hr> は単独で使われる命令で終了タグはありません。横線を引く働きをします。単独で使われるタグを「空の要素 (empty elements)」と呼びます。

　htmlは様々な命令を持っています。いくつか紹介しましょう。先ほどのファイルのbodyの部分を次のように書き換えて下さい。

<body>

<h1>レベル1の見出し</h1>

<h2>レベル2の見出し</h2>

<hr>

<p>ここに文章</p>

<ul>

<li> 番号なしリストです </li>

<li> unordered list </li>

<li> 3行目

</ul>

<hr>

<ol>

<li> 番号付きのリストです </li>

<li> ordered list </li>

<li> 3行目 </li>

</ol>

</body>

　htmlには「表を作る」「画像を表示する」「リストを作る」など様々なタグがあります。詳しくは

http://www.tohoho-web.com/www.htm　　とほほのWWW入門

などを参照してください。

　htmlのタグの中で一番重要なタグを紹介します。それはリンクを表すタグです。まず、もう1つファイルを作って下さい。ここでは例として z:\tmp\tmp2.html というファイルを作ります。ディレクトリ名やファイル名は皆さんの環境に応じて読み替えて下さい。次のような内容にしましょう。

<html>

<head>

<title>もう一つのファイル</title>

</head>

<body>

これはもう一つのファイルです。

</body>

</html>

　そして、先ほどの tmp1.html の body の部分を次のようにして下さい。

<body>

この文字の色は普通の色<br>

<a href="file:///z:\tmp\tmp2.html">ここを押す</a><br>

<a href="http://www.google.com/>ここを押す</a><br>

この文字の色も普通の色

</body>

　ここで <br> は単独で使われるタグであり、改行する命令です。「更新」ボタンを押して tmp1.html を再読み込みして下さい。<a> と </a> で囲まれた部分の文字の色は他の文字とは異なり、下線が引かれます（ブラウザの設定によっては異なるかも知れません）。そして、その文字をクリックすると href="...." で示されたファイルへジャンプします。これをハイパーリンクと呼びます。このハイパーリンクがhtml形式の一番大きな特徴であり、html形式のファイルをハイパーテキストと呼びます。

　リンクの上にマウスカーソルを持っていくと、ブラウザのステータスライン（一番下の行）にリンク先のurlが表示されます。

　ホームページビルダーあるいはDreamweaverなどのWebサイト作成ツールを使うと、htmlのタグについての知識がほとんどなくてもhtmlファイルを作成することが出来ます。また、Wordなどのワープロを使ってもhtmlファイルを作ることが出来ます。しかし、htmlの仕組みを理解した上で使うと、作業効率が更にあがるでしょう。htmlのハンドブックが色々と出版されているので、一冊手元に置いておくとよいでしょう。

課題1

　以下のような表を実現するためのhtmlのソースを示しなさい。

|  |  |
| --- | --- |
| 見出し1 | 見出し2 |
| 要素1 | 要素2 |
| 要素3 | 要素4 |

## スタイルシート

　前節でタグ <h1> を使った例を示しました。スタイルシートの機能を使うと、タブh1の見栄えをカスタマイズすることができます。<head> ブロックに以下のようなスタイルの定義を入れると、h1 の外見が変わります。

<style>

h1{

 color: darkblue;

 margin-left: 0px;

 margin-right: 0px;

 margin-top: 0px;

 margin-bottom: 20px;

 padding: 5px;

 width: 600px;

 border-style: solid;

 border-width: 0px 0px 2px 20px;

 border-color: #0000ff;

}

</style>

　この例ではhtmlファイルの中にスタイルを記述しましたが、<style> ～ </style> の部分を別ファイルに分離することができます。各々のhtmlファイルからスタイルを記述したファイルをインクルードすると、サイトに統一感ができます。mystyle.cssというファイルをインクルードするには、head ブロックに次のように書きます。

<link rel="stylesheet" type="text/css" href="mystyle.css">

　ファイルのある区間だけにスタイルを適用するには <div style="□□□">‥‥‥</div> と書きます。1つの段落だけにスタイルを適用するには <p style="□□□">‥‥‥</p> と書きます。

課題2

　ある段落に「文字色は赤」「背景色は黒」「行間は1.7倍」「横幅は500px」「中央寄せ」「周囲との間隔 (padding) は3em」というスタイルを設定したい。<p style="□□□">‥‥‥</p> と書くとき、□□□の部分に入る命令を書きなさい。

## 1つのページを構成する要素

　1つのページは1つのファイルで構成されているとは限りません。1つのページが複数のファイルから構成され、そのファイルが複数のサイトに分散している場合もあります。

課題3

　「引用を含むページ」をクリックして表示されるページに移動し、画面に書かれている問に答えなさい。

　この例のように、あるサイトのページの一部が別のサイトから引用されていることがあります。自分のサイトを作るときに別のサイトを引用する場合は、引用であることが明白であるようにしましょう。

## サーバ上で実行されるプログラム

　Webページが表示される仕組みは

1. ブラウザがサーバに対してファイルを要求する

2. サーバが要求されたファイルを送り返す

です。サーバ上のファイルは次の3種類に分類できます

1. 普通のファイル・・・ファイルをそのままクライアントに送ります。

2. phpファイル・・・指定されたファイルの一部にphpというプログラミング言語で書かれたプログラムが含まれています。phpで書かれた部分はプログラムの実行結果に置き換えられ、それ以外の部分はそのままクライアントに送ります。

3. cgi (common gateway interface) ファイル・・・このファイルはプログラムです。サーバ上で実行し、実行結果（プログラムが出力した文字列）をクライアントへ送ります。perl, ruby, pythonなどのプログラミング言語が用いられます。

　2.のファイルは拡張子にphpがついています。3. のファイルはサーバによって設定が異なるのですが、本実習で使用するサーバdenki.nara-edu.ac.jpは拡張子plを持つファイルをcgiファイルであると認識します。

課題4

(1)「phpの例（時刻表示）」と「その中身」をクリックし、「phpで書かれた部分」と「そのプログラムの実行結果」をそれぞれ示しなさい。IEの場合は「その中身」をクリックした後「ソース表示」で中身を閲覧できます。

(2)「cgiの例2（2の累乗）」と「その中身」をクリックし、表を生成している部分のperlのコードを抜き出しなさい。

## Formを用いたページ

　文字などを記入する欄があり、「送信」などのボタンを押すと情報がブラウザ側からサーバ側へ送られるページがあります。現在ではボタンをクリックするイベントに対応したJavaScriptが動作するケースが多いと思われますが、ここでは古典的な方法であるFormタグを使う方法を示します。

　まずは最もシンプルなフォームを見ましょう。「シンプルなフォーム：Get（php）」をクリックして下さい。htmlファイルのソースは以下のようになっています。

<html>

<head>

<meta http-equiv="content-type" content="text/html; charset=euc-jp">

<title>シンプルなフォーム</title>

</head>

<body>

<form method="get" action="action2g.php">

<p>

テキストを入れてください。

<input type="text" name="text1" size="20">

このテキストボックスの名前は text1 です。

<p>

<input type="submit" value="send">

</form>

</body>

</html>

　<form> </form> で囲まれた部分があり、その中に2つの要素があります。一つは <input type="text" でテキストボックスを表しています。名前はtext1です。もう一つは <input type="submit" で送信ボタンを表しています。送信ボタンは一種のリンク先を表しており、クリックすると <form action="action2g.php" で指定されたurlにジャンプします。マウスカーソルをボタンの上に置くと、ステータスバーにaction2g.php のurlが表示されます。クリックすると、以下のようなアクセスを行います（実際は HTTP/1.1 で通信を行いますが、ここでは簡略化して説明しています）。

GET /~yabu/edu/jyoho/html/action2g.php?text1=abcd HTTP/1.0

(空行)

　上の例はテキストボックスの中にabcdという文字列が入っている場合です。ファイル名の後に ? が付き、その後にデータが続いています。

　action2g.phpは送られてきたデータを処理する必要があるので、何らかのプログラムでないといけません。すなわちphpファイルかcgiファイルのいずれかです。action2g.phpの内容は以下の通りです。

<html>

<head>

<meta http-equiv="content-type" content="text/html; charset=euc-jp">

<title>Form の実験結果</title>

</head>

<body>

<p> 送られてきたデータは以下の通りです。

<p>

<?php

 foreach( $\_GET as $key => $value ){

 echo "$key = $value <br>";

 }

?>

</body>

</html>

　サーバ側で実行されるphpプログラムは、送られてきたデータを読みとります。GETメソッドでアクセスした場合、$\_GETという連想配列に内容が入ります。先ほどの例の場合は、変数 $\_GET{'text1'} の内容が 'abc' となります。

　このプログラムはクライアントから送られてきたデータを表示するだけですが、実用的には送られてきた内容をサーバ上のファイルに保存したり、あるいはデータベースサーバと通信してデータベースに格納したりします。

　サーバにデータを送る方法は2通りあります。1つはこの例で示したGETメソッドを使う方法で、もう1つはPOSTメソッドを使う方法です。先ほどの例はGETメソッドを使った例でした。以下のように使い分けます。

1. GETメソッドでデータを送る場合、データ量の制限があります。Webサーバによって制限値は異なりますがapacheのデフォルトは8192バイトです。大きなデータを送るときはPOSTメソッドを使います。

2. GETメソッドで送るとブラウザのurlの欄に送るデータが表示されます。すなわちデータがブラウザの履歴やサーバのログに残ります。従って、ユーザIDやパスワードを送るときにGETメソッドを使うのは良くありません。

課題5

　「3つの要素を持つフォーム Get（php）の内容は以下の通りです。

<body>

<form method="get" action="action.pl">

<p>

どちらかを選んで下さい。キーは radio1

<input type = "radio" name="radio1" value="val1" checked>1 つめの値

<input type = "radio" name="radio1" value="val2">2 つめの値

<p>

テキストを入れて下さい。キーは text1

<input type="text" name="text1" size="20">

<p>

この下はテキストエリアです。キーは text2<br>

<textarea name="text2" rows="3" cols="40">

Default text

</textarea>

<p>

<input type="submit" value="送信">

</body>

　このフォームにおいて「2つめの値」を選択し、テキストとして「abc」を入力し、「テキストエリアには「abc def（改行）xyz」と入れた場合、サーバへアクセスするときのurlを答えなさい。

　各フィールドを区切る文字と、キーと値を結ぶ文字には何が使われているか答えなさい。

## リンク元とサーバのログ

　ブラウザがWebサーバに対してファイルの送信を要求する場合、最小限で済ませるには

GET /~yabu/index.html HTTP/1.0

(空行)

でファイルを取得することが出来ます。しかし、IEなどのブラウザはそれ以外にも情報を送っています。「リンク元」「ブラウザの種類とバージョン名」などの情報です。例えば、次のように送っています。

GET / HTTP/1.1

Accept: image/gif, image/jpeg, image/pjpeg, image/pjpeg, application/x-shockwave-flash, application/vnd.ms-excel, application/vnd.ms-powerpoint, application/x-ms-application, application/x-ms-xbap, application/vnd.ms-xpsdocument, application/xaml+xml, application/msword, \*/\*

Referer: http://denki.nara-edu.ac.jp/~yabu/test.html

Accept-Language: ja

User-Agent: Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; YTB720; (R1 1.5); .NET CLR 2.0.50727; .NET CLR 3.0.4506.2152; .NET CLR 3.5.30729)

Accept-Encoding: gzip, deflate

Host: localhost

Connection: Keep-Alive

(空行)

　これまでの実習ではHTTPのプロトコルとしてバージョン1.0を使ってきましたが、IEやFirefoxなどのブラウザはバージョン1.1を使っています。バージョン1.1においてはGETの行に加えてHost: の行が必須となっています。

　Referer: がリンク元のurlを示し、User-Agent: がブラウザの種類のバージョンです。

　これらの情報はサーバのログに記録されます。サーバ用ソフトウェアとして一番メジャーなソフトであるapacheのログの例を示します。1回のアクセスのログは次のようになります。

192.168.xx.yy - - [29/Oct/2015:15:50:49 +0900] "GET /~yabu/soft/original/ HTTP/1.1" 200 13797 "http://www.google.co.jp/url?sa=t&rct=j&q=&esrc=s&source=web&cd=9&ved=0CFUQFjAIahUKEwj3-8HPi-fIAhWFE5QKHWu9BCk&url=http%3A%2F%2Fdenki.nara-edu.ac.jp%2F~yabu%2Fsoft%2Foriginal%2F&usg=AFQjCNFVU0OunVg-wEcpcvyIgm5UoEXv0Q" "Mozilla/5.0 (Windows NT 6.1; WOW64; Trident/7.0; rv:11.0) like Gecko"

　この例は、googleで何らかの検索をした結果、このサイトに来たことを示しています。

　また、サーバとクライアントはネットワーク通信を行いますので、サーバはクライアントのIPアドレスを知っています。クライアントのIPアドレス（この例では 192.168.xx.yy）あるいはそのIPアドレスに対応するホスト名もログに記録されます。自宅や携帯電話からアクセスする場合、プロバイダやキャリアが個人情報を開示すれば、誰かを特定することが出来ます。

## クッキー（cookie）とは

　クッキーといえば食べるお菓子ですが、ここで取り上げるクッキーはWebサーバとWebブラウザの間でやりとりされます。まずは次の図をご覧下さい。



図1　クッキーの動作

　図のように1回目のアクセスの時、サーバからクライアント（ブラウザ）へクッキーをセットせよという命令が送られます。具体的には次のような行です。

Set-Cookie: ans=yes; expires=......; path=/; domain=denki.nara-edu.ac.jp

　この例ではansという変数を設定し、その値をyesに設定しています。expires= の部分はそのクッキーの有効期限であり「○○年○月○日○時○分○秒」を表す値がエンコードされています。

　ブラウザ側ではこのクッキーの値を保持しておき、「domainが後方一致するサーバ」の「pathが前方一致するフォルダより下にあるファイル」をgetするとき、保持していたクッキーの値を送ります。

　クライアントがアクセスするページがphpファイルあるいはcgiファイルの場合、phpやcgiのプログラムは送られてきたリクエストに含まれているcookieの値を読みとることが出来ます。これによりクライアントの情報（例えば前回いつ訪れたか）などの情報をサーバ側で得ることが出来ます。

　今、Webブラウザがどのようなcookieを保持しているかは次のように操作すると見ることが出来ます。

* IE 11の場合、「ツール」→「インターネットオプション」→「全般」の閲覧の履歴の「設定」→「ファイルの表示」でファイルを表示するウィンドウを開くと、その中に「cookie:yabu@aaa.bbb.com」のようなファイル名でCookieが保存されています。秀丸エディタを起動し、そのファイルを秀丸エディタにドラッグして開いて下さい。1行目が変数の名前、2行目が変数の値、3行目がそのクッキーを送ってきたサイトです。クッキーが複数ある場合は「\*」が区切りです。Cookieを削除するには「閲覧の履歴」の「削除」ボタンをクリックし、「cookieを削除」ボタンを押して下さい。
* Firefox 33.1.1の場合、「ツール」→「オプション」→「プライバシー」→「Cookieを個別に削除」で見ることが出来ます。

　これを見ると、知らない間に大量のCookieを受け入れていたことが分かります。

課題6

　クッキーの値が確認しやすいのはFirefoxなので、この課題はFirefoxを使って行って下さい。

　次のサイトにアクセスしなさい。送られてきたクッキーの名前、値、有効期限などの情報を調べなさい。一旦、自分が保持しているCookieを全て削除してから実行します。それによって、パスワードの入力を省略出来ていたサイトにおいて、パスワードを再入力する必要が生じたり、amazonなどで過去の閲覧履歴が使えなくなるかも知れませんが、それは諦めて下さい。

* http://denki.nara-edu.ac.jp/~yabu/edu/jyoho/html/cookie-send2.php
* http://www.yahoo.co.jp
* http://www.google.co.jp

www.yahoo.co.jpとwww.google.co.jpにおいては何か検索を行いなさい。

## ブラウザ側で実行されるプログラム

　htmlファイルの中にJavaScript（プログラミング言語の名前）で書かれたプログラムが含まれている場合があります。そのようなページを表示するとき、ブラウザはJavaScriptを実行します。その結果、動的なページ（見る人、見るタイミング、ユーザの操作などによって異なるページになること）になります。

　まず、「JavaScriptの例（2の累乗）」を表示して下さい。「ソースを表示」でhtmlファイルの内容を見ることができます。JavaScriptのプログラムが表の部分を生成していることが分かります。

課題7

(1) 2の累乗の表を生成しているプログラム部分を抜き出しなさい。

(2) 2の累乗を表示するページはphpのページとJavaScriptのページの2通りがあったが、結果は似ていた。しかし、そのページが生成される方法は全く異なる。どのように異なるのか、説明しなさい。

　Webブラウザ上におけるクリックやマウスの移動などのイベントに対応してJavaScriptが動作し、ページの内容を動的に書き換えることができます。「JavaScriptの例（ボタンクリックのイベント）」「JavaScriptの例（マウスのイベント）」のページを表示してみなさい。

課題8

(1)「JavaScriptの例（マウスのイベント）」を開いて下さい。「表示」→「ソース」で内容を見ることができます。「鹿」「夕暮れ」の文字に設定されているタグを答えなさい。またそれらの文字の上にマウスカーソルが来たときにコールされる関数名を答えなさい。

(2)「JavaScriptの例（ポップアップウィンドウもどき）」を開いて下さい。ウィンドウもどきが移動します。ソースを表示し、「この部分が移動します」という文字列に設定されているタグを答えなさい。

　JavaScriptがサーバと通信して動的にページを書き換えることができます。これはAjaxと呼ばれる技術です。Googleの検索ウィンドウに文字を入力すると、1文字入力するたびに候補を表示します。これは1文字入力するたびにサーバと通信して候補を取得しているからです。Googleの地図機能もAjaxを利用しています。

　「JavaScriptの例（サーバとの通信　同期）」を開きなさい。テキストボックスで「キーを押した」というイベントが発生するたびに、テキストボックスの内容をサーバに送信します。サーバは送られて来た文字列をそのまま返します。

課題9

　上記のページはテキストボックスに1文字入力するたびに、あるurlにアクセスする。そのurlを答えなさい。