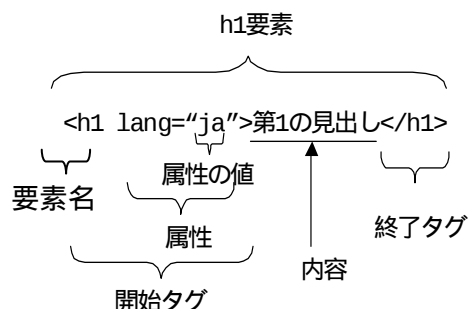


第 10 回：外国語ウェブページの作成

2. テキスト文書としてのウェブページ (前回からの続き)

2.2. HTML 文書の基本

- HTML 文書では、文書の構造を「マークアップ」markup というきめられた方法で表現する。マークアップは、「タグ」と呼ばれる半角の三角カッコで囲まれた文字列 (例：<h1>) によっておこなう。マークアップ各部の名称は、以下のとおり。

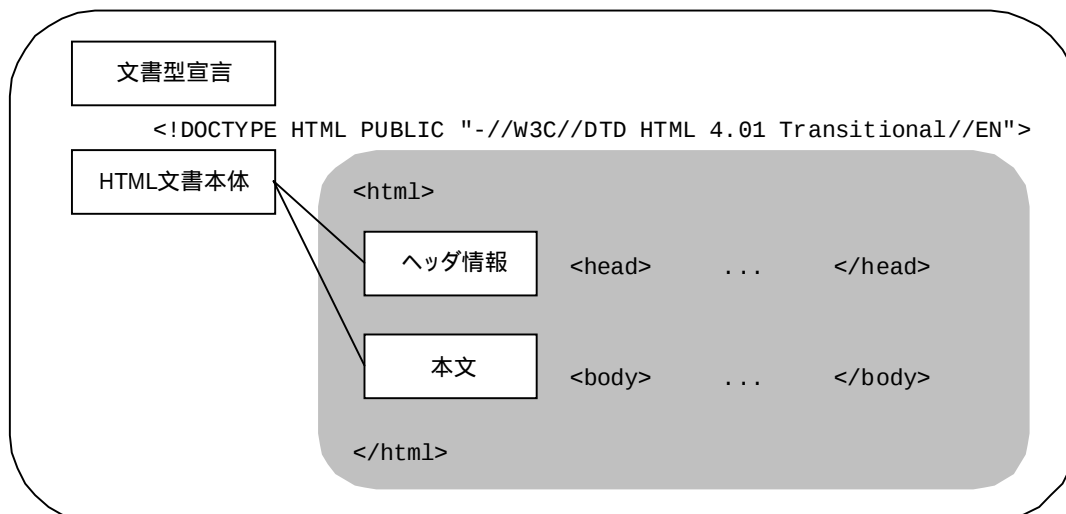


タグの内容 (要素名, 属性名, 図では h1 と lang) は必ず半角 (ASCII) で記述すること (大文字, 小文字どちらでも OK。おススメは全て小文字で入力する方法)。

- 4つの記号, 角括弧 < と >, アンパサンド &, 2重引用符 " は HTML 文書のマークアップに使う特別な記号である。これらを文章中の文字として使う場合は, マークアップの一部と混同されないよう, その文字を「文字実体参照 (Character Entity References)」と呼ばれる方法で置き換えて表現する必要がある。

<	< (less than)	&	& (ampersand)
>	> (greater than)	"	" (quotation)

- テキストファイルを閲覧する場合と異なり, HTML 文書では, **余分な空白(半角スペースやタブ)と改行は無視される**。逆に言うと, 改行や空白は, 作成者がテキストを見やすくするために適宜入れてよい (全角スペースはそのまま表示されるので注意)。
- HTML 文書の基本構造は, 次のとおり:



文書型宣言は, 文書先頭に記述し, その文書が HTML のどのバージョンに従

っているかを示す (HTML にはバージョンが複数ある)。

「ヘッダ情報」は、タイトル要素 <title> ... </title> のほか、文字エンコード方式情報 (後述) など、文書に関する説明(「メタ情報」という)を入れる。

- HTML についての情報
 - ◇ HTML 4.01 仕様書 (Specification): <http://www.w3.org/TR/html4/>
 - ◇ HTML 文法チェッカー Another HTML-lint gateway :
<http://openlab.ring.gr.jp/k16/htmlhint/htmlhint.html>
 - ◇ 情報システムセンター「初めての HTML 改定版 ver.1.2」:
<http://www.cs.reitaku-u.ac.jp/infosci/netwk/html-v12.html>
 - ◇ その他, 参考になる無数のウェブページや書籍がある。活用しよう!
 - ◇ よいデザインのページに出会ったら, ソースを分析して勉強しよう。
Internet Explorer : [表示]→[ソース](または右クリックで「ソースの表示」)
Opera : [表示]→[ページのソースを表示]
Netscape Communicator : [表示]→[ページのソース]
 - ◇ HTML 文書の作成については, 後期の情報処理演習 B の授業で本格的に学ぶので, 興味のある人はぜひ履修してほしい。

2.3. 多言語 HTML 文書とエンコード

これまで学習したように, テキスト文書は言語に合わせて正しい文字エンコード方式を指定し, 保存する必要がある。外国語で HTML 文書を作成する場合も, テキストを正しいエンコード方式で処理する必要がある。ファイルの保存・再編集の際は充分注意しよう。

各言語・地域で標準的に使われるエンコード方式とコードページ (第 6 回資料より)

言語(地域)	標準的な文字エンコードの名称	Windows での名称	Windows コードページ Codepage
日本語	Shift JIS	日本語 (シフト JIS)	932
西ヨーロッパ言語	ISO-8859-1	西ヨーロッパ言語(ISO)	
		西ヨーロッパ言語(Windows)	1252
中国語 (中国)	GB2312	簡体字中国語 (GB2312)	936
中国語 (台湾)	BIG5	繁体字中国語 (Big5)	950
韓国語	EUC-KR	韓国語 (EUC)	
		韓国語	949
タイ語	TIS-620		
		タイ語 (Windows)	874
多言語	UTF-8	Unicode (UTF-8)	

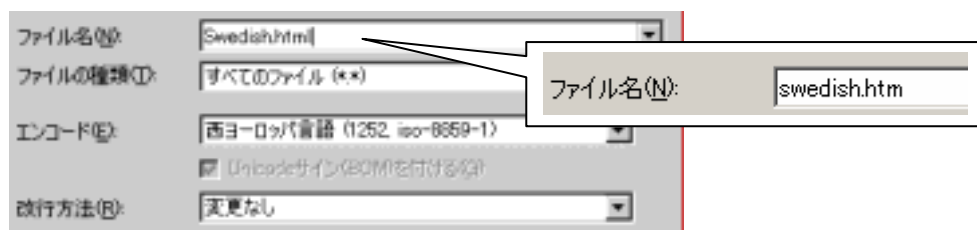
HTML 文書では, < や > , " など, タグ情報は全て ASCII に含まれる文字で記述する。Unicode は全ての文字を 16 ビットで表現するため, Unicode (UTF-16) で書かれた HTML 文書を表示するためには Web サーバやブラウザが ASCII と Unicode の違いを認識してタグを読み分けることになり, 大変不便である。そこで, HTML 文書を Unicode を用いて作成する場合はエンコード方式として UTF-8 を使用するのが普通である。

3. 外国語ウェブページの作成

3.1. EmEditor での外国語 HTML 文書の編集

- ファイルを保存: メニューバーの [ファイル] → [名前をつけて保存] を選択し, ファイルの種類を「全てのファイル」(または「HTML」)とし, 「ファイル名」には拡張子

を含む正確なファイル名 (HTML 文書の拡張子は **.html** または **.htm**) をつける (下図)。「エンコード」も言語にあわせて変更する。多言語テキストを Unicode で保存する場合には、「エンコード」に「UTF-8」を指定する。



- ファイルを開く：メニューバーの [ファイル] → [開く] を選択し、「ファイルの種類」を「HTML (*.htm, *.html ...)」 (または「全てのファイル」) に指定し、HTML 文書を選択する (下図、「Text (*.txt)」のままで HTML 文書が選択できないので注意)。文字エンコード方式の設定は「コードページ」でおこなうか、ファイルを開いた後で「ファイル」→「コードページを変更して読み直し」を選択し、エンコード方式を設定する。

「ファイルの種類」をいちいち変更しなければならない上記の方法がわずらわしい場合、ショートカットを利用するとよい。開きたいファイルのアイコン上でマウスを右クリックし、[送る] → [EmEditor] を選択すると、すばやくファイルを開くことができる (右図参照、自宅 PC にインストールした EmEditor の場合には、右クリックしたメニューに「EmEditor」という項目ができる)。この場合、文字エンコード方式は「ファイル」メニューの「コードページを変更して読み直し」を用いてファイルを開いてからおこなう。



なお、テキストの拡張子が htm ないし html のファイルを開くと、EmEditor の編集モードが自動的に切り替わり、タグが色分け表示される (見やすい!)。

3.2 Word での外国語 HTML 文書の編集

Word も多言語テキストの編集に対応しているので、多言語 HTML 文書を編集することができる。ただし、以下のように HTML 文書をテキスト文書として開くための設定を必ずしておく必要があるので注意してほしい。

- **【重要！】** HTML 文書を編集する前に、[ツール]から[オプション]を開き、「全般」タブの「文書を開くときにファイル形式を確認する」をチェックし、テキスト文書として HTML を読み込めるように設定しておく。大学の PC ではログオンごとに設定が必要なので、充分注意すること！。
- ファイルを保存：ファイルを開く際、ファイルの種類を「エンコードされたテキスト」にする (右図参照、「HTML 文書」にはしないこと)。ファイル名は拡張子 **html** (または **htm**) までつける。「エンコードされたテキスト」として開き、「ファイルの変換」ダイアログで、エンコードを指定する。これで、HTML 文書をテキストとして編集できるようになる。



注意： Word には、保存するファイルの形式として「HTML 文書」があるが、この形式は外国語の HTML 文書を作成する場合には**使ってはならない**。Word 文書の書式維持のため独自のタグが挿入されてしまい、Word 以外のアプリケーションで適切な編集をすることがで

きなくなるばかりか、外国語の文字エンコード方式でソースが保存されない。



- ファイルを開く：ファイルの種類を「すべてのファイル (*.*)」にする。(上の「準備」で記したオプションの設定が出来ていないと、HTML 文書を選択すると Word の HTML 文書編集モードで開いてしまい、HTML のタグや文字の情報が勝手に修正されてしまう。この場合、保存せずに終了させ、設定を変更すること。)

3.3. HTML 文書のファイル名のつけかたに関する注意

- HTML 文書のファイル名には、ASCII の基本アルファベットと英数字、ハイフン、アンダーバーのみを使い、日本語や、各言語の特殊なアルファベットや記号、スペースは使わないこと。なお、ファイル名
- インターネット上で公開する場合、ファイル名の大文字・小文字は厳密に区別されるので注意。index.htm と index.html, Index.htm は全て異なるファイル名である。(Windows 上では大文字小文字の違いは区別されないので、1 番目と 3 番目の違いに気づかないことがある！)

実習 5：【この実習は先週既に行っている人もいる】 **English.txt** を EmEditor で開き、「西ヨーロッパ言語」をエンコード方式に指定してポンド記号などを正しく表示しなさい。「名前をつけて保存」を選択し、テキストを **English.html** というファイル名で新たに保存しなさい (コードページは「西ヨーロッパ言語」としなさい)。アイコンをダブルクリックして WWW ブラウザで開いて表示を確認し、必要ならばエンコードを正しく設定しなさい。(続いて実習 6.をおこないなさい。)

実習 6：別紙のサンプルを参考にしながら、EmEditor を使って **English.html** に HTML のマークアップを行いなさい。主な編集のポイントを以下にリストする。

- 文書の基本構造をあらわす html, head, body 要素を入れる。
- ヘッダ要素：title 要素の内容を「Editorial Note of Barnes (1989)」とする。
- h1 (大見出し)要素を「Editorial Note」とし、本文冒頭に置く。h2 (中見出し)は「Julian Barnes (1989) A History...」にする。
- Paperback Version..., Price ..., ISBN ..., First published... を番号なしリストにする：


```
<ul>
        <li>Paperback ... </li>
        <li>Price... </li>
        <li>ISBN ... </li>
        <li>First published... </li>
      </ul>
```
- Editorial Note の本文を一つの段落を表す <p> ... </p> で囲む。
- 文書の末尾に address 要素 (<address> ... </address>) を加え、名前(アルファベット)と学籍番号、メールアドレスを記入する (外国語のテキストであるので、名前に漢字は使わず、ASCII で入力すること)。
- 書名 (A History of ..., Metroland, Before She..., Flaubert's ... etc.) は斜字体 (<i> ... </i>で囲む) にする。
- h1 要素にある作家名に、以下のように本人のウェブページへのリンクを作成する (<a> タグの href=" ... " の中にウェブページの URL を記述する。終了

タグ `</a>` を忘れずに)。

```
<h1><a href="http://www.julianbarnes.com">Julian Barnes</a> (1989)
<i>A History of ... </i></h1>
```

編集後ファイルを上書き保存し、WWW ブラウザで再度表示し (ツールバーの「更新」ボタンをクリックする)、別紙サンプルのように表示されるか確認しなさい。



実習 7【Unicode を使った HTML 文書の作成】: 自己紹介の ウェブページを作成しなさい。第 5 回の課題として外国語で作成した自己紹介のテキスト intro.txt を [あなたのユーザ ID].html (←四角カッコは不要) というファイル名で保存しなさい (intro.txt は [No5] フォルダに保存されているはず。新たに作成する HTML 文書の保存場所は自由)。文字エンコード方式は intro.txt と同様、utf-8 (BOM なし) に設定しなさい。実習 6 のサンプルや、以下の指示を参照しながら HTML に従ってマークアップをおこないなさい (ユーザ ID が schiba の場合、ファイル名は **schiba.html** となる)。以下は、作成のヒント:

- 文書の基本構造をあらわす html, head, body 要素を必ず入れる。
- ヘッダ要素: title 要素の内容を「Introducing ○○ in /Japanese」とする (○○には自分の名前を ASCII の基本アルファベットで記す。には言語名を英語で記す)。

Introducing CHIBA Shoji in German/Japanese

↑	↑								
自分の名前を ASCII で	テキストの言語名を英語で								
	<table border="1"> <tr> <td>English</td> <td>Thai</td> </tr> <tr> <td>German</td> <td>Simplified Chinese</td> </tr> <tr> <td>Spanish</td> <td>Traditional Chinese</td> </tr> <tr> <td>French</td> <td>Korean</td> </tr> </table>	English	Thai	German	Simplified Chinese	Spanish	Traditional Chinese	French	Korean
English	Thai								
German	Simplified Chinese								
Spanish	Traditional Chinese								
French	Korean								

実際のソースは、以下ようになる:

```
<head>
<title>Introducing CHIBA Shoji in German/Japanese</title>
</head>
```

title 要素は head 要素の中に「入れ子」にすることに注意。</head>は必須!

- ページの大見出しに h1 要素を使い、本文冒頭に置く (内容、言語は自由)。
- 本文の内容にあわせ、段落を表す p 要素や中見出しの h2、箇条書きを表す ul, li 要素などを使う (</p> や , といった終了タグを忘れずに)。
- 文書の末尾に address 要素 (<address> ... </address>) を加え、名前(アルファベット)と学籍番号、メールアドレスを記入する (外国語のテキストであるので、名前に漢字は使わず、ASCII の基本アルファベットで入力すること)。
- 外国語版のページはその外国語の文字エンコード方式で、日本語版は Shift_JIS にて保存し、エンコードを確認して編集すること。

外国語のウェブページをブラウザで開く際におこる文字化けは、エンコードの再指定で直るが、一般にわずらわしい作業である。本格的に外国語ウェブページを作る際に覚えておきたい、文字化けを防ぐ方法については次回授業で取り上げる。

自己紹介のページは、次回授業で説明する文字エンコードや言語の記述を行い、次回授業翌日を締め切りに提出してもらう予定である。内容も評価に加える予定なので、趣味など、関連する内容をリンク集にまとめたり、画像を用意するなど、各自工夫して自己紹介のコンテンツを充実させておくこと。なお、自己紹介のテキストは 2 つの言語で厳密に同じものでなくともかまわない (紹介文の中でもう一方の言語の内容について簡単に紹介するとよい)。

(以上)