

多言語ウェブページの作成 (1)

1. テキスト閲覧環境としてのウェブブラウザ

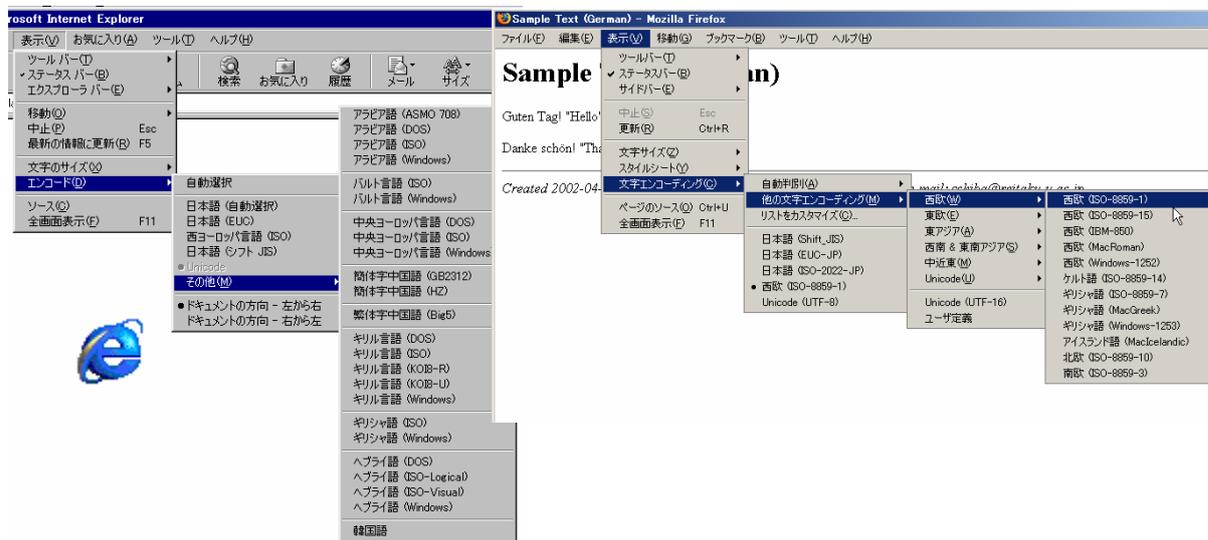
1.1. 外国語テキストの閲覧

これまで授業で見てきたように、テキスト文書はさまざまなソフトウェアで利用できる。外国語のエンコード方式のテキスト文書を閲覧 (表示, 印刷) できるソフトウェアとして最も手軽に使えるのは WWW ブラウザである。ブラウザは、テキストを読み込み、指定されたエンコード方式に応じて文字を適切なフォントで表示する。

外国語の文字エンコード方式で作成されたテキスト文書は、ウェブページを閲覧する場合と同様の方法でエンコードを指定し、ブラウザ上に正しく表示させることができる。手順は以下のとおり。

1. ブラウザを起動する
2. マイコンピュータなどでテキストファイルの入っているフォルダを開く
3. テキストファイルのアイコンをブラウザにドラッグ&ドロップ (マウスでアイコンをドラッグしてブラウザ上にもっていき、マウスボタンを放す)
 - ※ ブラウザの[ファイル]→[開く]でテキスト文書を指定すると、ブラウザでなくテキスト文書が関連付けられているソフトウェアが起動することがある。
4. テキストがブラウザで開かれる
5. 文字エンコード方式を選択し、文字を正しく表示する

※ Internet Explorer と Mozilla FireFox の文字エンコード方式の選択メニュー



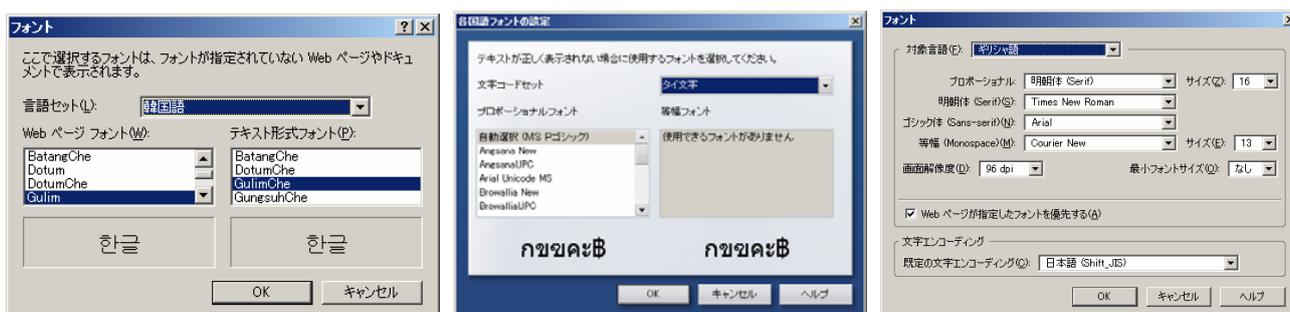
文字エンコード方式の表記はブラウザにより若干異なる (e.g. 「繁体字中国語」⇔「中国語」; 「西ヨーロッパ言語」⇔「欧米」)。また、Internet Explorer や Opera では、「西ヨーロッパ言語 (ISO)」といった標準的なエンコード方式に加え、「西ヨーロッパ言語 (Windows) (Windows コードページ 1252) のような Windows 独自の拡張を含むエンコード方式が利用できる。

実習1: Kadai サーバの [schiba]→[2006f1] にある [No9] フォルダを file_server の Home にコピーしなさい。Internet Explorer を使い、コピーした [No9] フォルダにある English.txt (英語), Korean.txt (韓国語), SimplifiedChinese.txt (簡体字中国語), TraditionalChinese.txt (繁体字中国語) を表示させ、ブラウザのメニューバーの「表示」→「エンコード」で正しいエンコードを指定し、文字を正しく表示させなさい (文字エンコード方式としてそれぞれ「西ヨーロッパ言語(ISO)」「韓国語」「簡体字中国語 (GB2312)」「(繁体字)中国語 (Big5)」を選ぶ)。同様の作業を Opera (エンコード方式は、[表示]→[エンコード] から選ぶ), Mozilla FireFox でも行いなさい (エンコード方式は [表示]→[文字エンコーディング] から選ぶ)。

1.2. フォントの選択とフォントの種類

Internet Explorer や Opera, Mozilla Firefox など、多くのブラウザでは、エンコードとともに言語・地域のフォントが切り替わるよう言語ごとにフォントが設定できる。

実習2: 自分が利用する外国語について、Internet Explorer や Opera, Mozilla Firefox のフォントの設定を確認しなさい。Internet Explorer ではブラウザのメニューバーの「ツール」→「インターネットオプション」の「全般」タブの「フォント」を、Opera は「ツール」→「設定」の「詳細設定」タブの「フォント」を、Mozilla FireFox は「ツール」→「オプション」の「コンテンツ」タブにある「フォントと配色」の「詳細設定」を開く。



各言語・地域の文字エンコード方式ごとに、「プロポーショナルフォント」と「等幅フォント」が設定されていることがわかる。

2. テキスト文書としてのウェブページ

2.1. HTML 文書と拡張子

- ウェブページは、HTML (HyperText Markup Language) という決まりに従って書かれている。このため、ウェブページのことを HTML 文書と呼ぶ。
- HTML 文書の中身(「ソース」という)は、実はテキスト文書である。つまり、エンコード方法さえおさえれば、どんな言語のウェブページでもテキストとまったく同じように編集できる。
- Windows では、ファイルの種類をファイル名の末尾の記号で示している。この記号を **拡張子 file extension** という。テキスト文書でよく使われるファイルの拡張子は txt だが、HTML 文書では htm または html が用いられることが多い(おすすめは **html**)。、
- Windows は拡張子によってどのアプリケーションで開くかをあらかじめ登録し、起動しやすくしている。この対応を **関連付け association** といい、関連付けられたファイルは、「マイコンピュータ」からファイルをダブルクリックするだけで開くことができる。大学の PC では htm や html という拡張子がついたファイルは Internet Explorer に関連づけられている。

Windows の初期設定では、関連付けられているファイルの拡張子は省略される。拡張子を含んだ「正確な」ファイル名を知らないと、HTML 文書を編集する際にトラブルが起こることがある。拡張子を含むファイル名を表示する方法を覚えておこう。

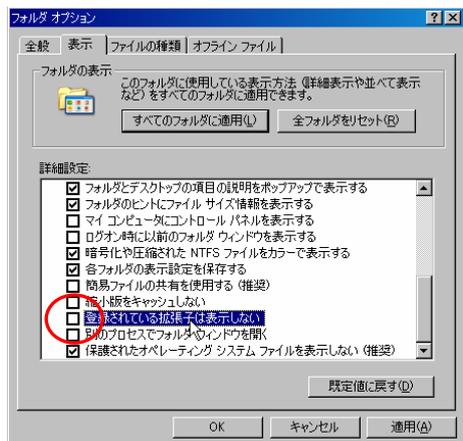
【復習】実習 3 : 「マイコンピュータ」にコピーした [No9] フォルダにある file という名前の 3 つのファイルを比べなさい。ログオンした時点では「マイコンピュータ」でファイルを一覧表示すると、これらのファイルは全て file という名前で表示される。下記の手順で設定を変更して拡張子を表示させ、正しいファイル名と拡張子名を書き出しなさい。



ファイルの種類	正確なファイル名	拡張子
Microsoft Word 文書		
テキスト文書(ドキュメント)		
Excel ワークシート		

1. 「マイコンピュータ」を開き、メニューバーの[ツール]から[フォルダオプション]を開く。
2. 「表示」タブを選択する。
3. 「登録されている拡張子は表示しない」という項目のチェックをはずす(右図参照)。
4. 「OK」ボタンを押す。これで、拡張子付きのファイル名が表示されるようになる。

※この設定は、大学のパソコンではログオンする度におこなわなければならない。自分のパソコンの場合、一度設定すれば同じ設定が適用される。



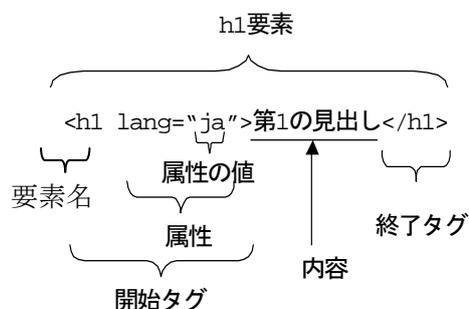
実習 4 : English.txt を EmEditor で開き、「西ヨーロッパ言語」を文字エンコード方式に指定してポンド記号などを正しく表示しなさい。一旦 EmEditor を閉じ、ファイル名の拡張子を.txt から.html に変更しなさい(実習 3 の拡張子の表示設定を必ず変更しておくこと)。ファイルアイコンをダブルクリックして、ファイルが WWW ブラウザ Internet Explorer で開き、「コードページ」を「西ヨーロッパ言語」としてテキストが正しく表示されることを確認しなさい。

- ウェブページの拡張子 .html (または .htm) は、大学 PC では Internet Explorer に関連付けられており、ファイルアイコンをダブルクリックすると Internet Explorer が起動する。EmEditor で内容を編集したり、Internet Explorer 以外の WWW ブラウザでファイルを開く場合はソフトウェアからファイルを指定して開く操作をおこなう。EmEditor の場合は、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「全てのファイル」(または「HTML」) に指定して HTML 文書を選択する(「Text (*.txt)」のままでは HTML 文書が選択できない。もっと手軽な修正方法については後で紹介する)。

2. テキスト文書としてのウェブページ

2.1. 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。

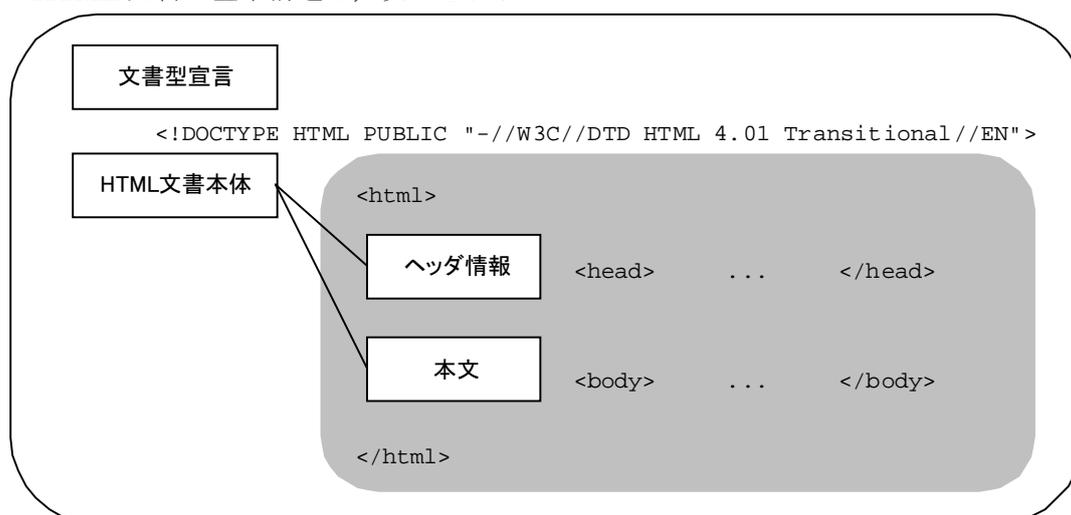


タグの内容(要素名, 属性名, 図では h1 と lang) は必ず半角(ASCII)で記述すること(大文字, 小文字どちらでもOK。おすすめは全て小文字で入力する方法)。

- 4つの記号, 角括弧 < と >, アンパーサンド &, 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/htmlint/htmlint.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。

2.2. HTML 文書と文字エンコード方式

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)

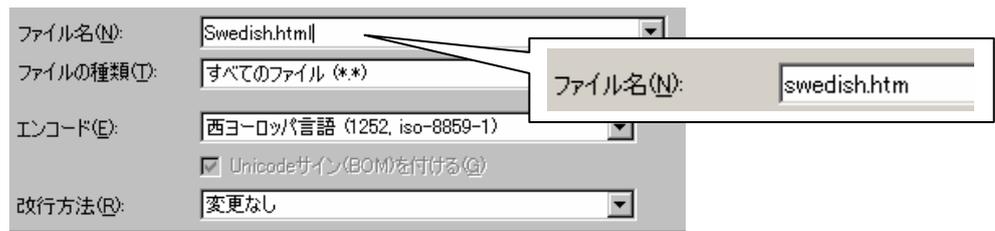
言語(地域)	標準的な文字エンコードの名称	Windows での名称	Windows コードページ Codepage
日本語	Shift JIS	日本語 (シフト JIS)	932
西ヨーロッパ言語	ISO-8859-1	西ヨーロッパ言語(ISO)	
		西ヨーロッパ言語(Windows)	1252
中国語 (中国)	GB2312	簡体字中国語 (GB2312)	936
中国語 (台湾)	BIG5	繁体字中国語 (Big5)	950
韓国語	EUC-KR	韓国語 (EUC)	
		韓国語	949
タイ語	TIS-620		
		タイ語 (Windows)	874
多言語	UTF-8	Unicode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

実習 4：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 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。



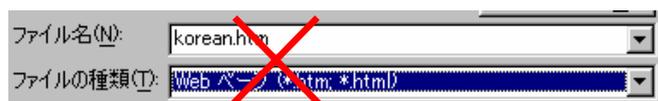
3.2 Word での外国語 HTML 文書の編集

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くなる。



- 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！)

サンプルページ : English.html

Editorial Note

Julian Barnes (1989) *A History of the World in 10 1/2 Chapters*

- Paperback Version published by Picador, London.
- Price: UK £6.99, EUR 12.23.
- ISBN 0 330 31399 1
- First published 1989 by Jonathan Cape

Julian Barnes has published seven novels, *Metroland*, *Before She Met Me*, *Flaubert's Parrot*, *Staring at the Sun*, *A History of the World in 10 1/2 Chapters*, *Talking it Over*, and *The Porcupine*, and a collection of journalism, *Letters From London 1990-1995*: his work has been translated into more than twenty languages. In France he is the only writer to have won both the Prix Médicis (for *Flaubert's Parrot*) and the Prix Fémina (for *Talking it Over*). In 1993 he was awarded the Shakespeare Prize by the FVS Foundation of Hamburg.

CHIBA Shoju, 0000000000, schiba@reitaku-u.ac.jp

サンプルページのソース (改行記号を ↓ で示しているが、従わなくともよい)