

## 第 8 回：外国語ウェブページの作成

### 1. テキスト閲覧環境としてのウェブブラウザ

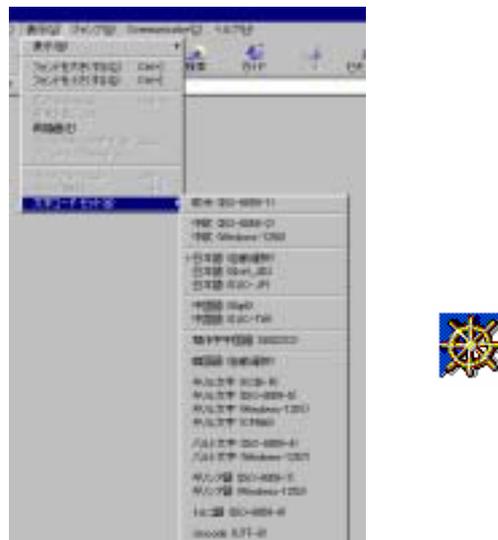
#### 1.1. 外国語テキストの閲覧

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 でエンコード方式を選択      ※ Netscape でエンコード方式を選択



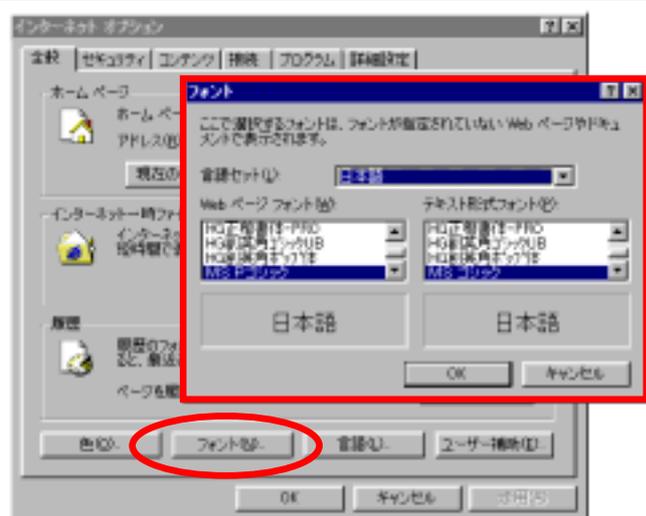
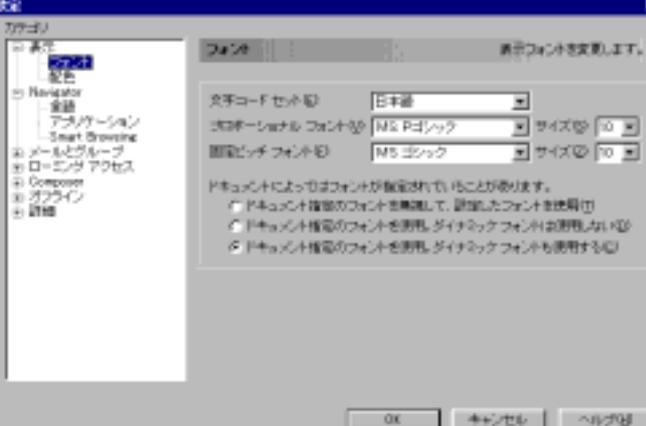
文字エンコード方式の表記はブラウザにより若干異なる (e.g. 「繁体字中国語」⇔「中国語」; 「西ヨーロッパ言語」⇔「欧米」)。また、Internet Explorer や Opera では、「西ヨーロッパ言語 (ISO)」といった標準的なエンコード方式に加え、「西ヨーロッパ言語 (Windows) (Windows コードページ 1252) のような Windows 独自の拡張を含むエンコード方式が利用できる。

実習 1 : Kadai サーバの [schiba]→[2004f1] にある [No8] 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 [No8] 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), Netscape Communicator でも行いなさい (エンコード方式は [表示]→[文字コードセット] から選ぶ)。

### 1.2. フォントの選択とフォントの種類

Internet Explorer は、特に設定しなくとも、標準の状態では各国語のフォントを使えるが、Netscape Communicator の場合は、ソフトウェアのフォントの設定が不完全なため、言語によっては、ユーザがフォントを設定してやらなければテキストを正しく表示できないことがある。実習 1 の中国語テキストを Netscape Navigator で閲覧しようとするとう文字化けが起こるが、Netscape 自体は文字エンコードをきちんと解釈するので、フォントさえ正しく指定すれば、表示ができるようになる。

Internet Explorer, Netscape でのフォントの設定・確認の仕方は以下のとおり :

<p><b>Internet Explorer :</b></p> <ol style="list-style-type: none"> <li>1. [ツール]→[インターネットオプション]を開く。「全般」タブの下にある「フォント」ボタンを押して、「フォント」画面を開く。</li> <li>2. 「言語セット」を選び、言語セットごとにフォントを選択する。</li> <li>3. [OK]ボタンを押す。</li> </ol>	
<p><b>Netscape Communicator :</b></p> <ol style="list-style-type: none"> <li>1. [編集]→[設定]で、「設定」画面を開く。</li> <li>2. 「表示」カテゴリから「フォント」をマウスで選択する。</li> <li>3. 文字コードセットごとにフォントを選択する。</li> <li>4. [OK]ボタンを押す。</li> </ol>	

Internet Explorer, Netscape Communicator とも、各言語・地域の文字エンコード方式(呼び名はソフトによって異なる)ごとにプロポーショナルフォント, 等幅フォントをそれぞれ指定することができる (フォントの種類については第 4 回資料を参照)。

実習 2: 下記フォントリストや Internet Explorer の設定を参考に, Netscape Communicator で韓国語, 中国語(簡体字, 繁体字)が表示できるようにフォントを設定しなさい。

WindowsXP 日本語版に搭載されている韓国語, 中国語フォントの例 (第 4 回資料より)  
韓国語

Batang	안녕하십니까? 'Hello!'	プロポーショナル, セリフ
BatanChe	안녕하십니까? 'Hello!'	等幅, セリフ
Dotum	안녕하십니까? 'Hello!'	プロポーショナル, サンセリフ
DotumChe	안녕하십니까? 'Hello!'	等幅

中国語 (簡体字)

SimSun	我喜欢学习中文。I wanna study Chinese.	等幅
SimHei	我喜欢学习中文。I wanna study Chinese.	等幅

中国語 (繁体字)

PmingLiU	我喜欢学习中文。I wanna study Chise.	プロポーショナル
MingLiU	我喜欢学习中文。I wanna study Chinese.	等幅

実習 3: Netscape Communicator で Yahoo! の各国版を閲覧しなさい。アメリカの本家のページ <http://www.yahoo.com> の下方に各国語版へのリンクがあるので参照し, 自分の言語のページへすすみなさい。うまく閲覧できたら, サイトをブックマークに登録しなさい。

大学の PC の仕様では, Internet Explorer の設定に施した変更は, ログオフするとリセットされてしまい, 保存されない。これに対し, Opera や Netscape Communicator では, 各ユーザの file\_server の Home (F:) の「Opera7」「netscape」フォルダの中に設定内容が保存される。そのため, Netscape でおこなった各種設定 (ブックマークやフォントの指定) はログオフ後も残る。

大学の Netscape のバージョンはかなり古い (4.72, 現在のバージョンは 7.1) ので使い勝手はあまりよくない。また, ロシア語などに使われるキリル文字の表示や, スタイルシートの表現に問題があるなど, いくつか重要なバグが残っている。よって, 大学で WWW を上手に使いこなすためのツールとしては Opera が最適, と言えるかもしれない。

## 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。

実習 4 : 「マイコンピュータ」にコピーした [No8] 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さなさい。



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		

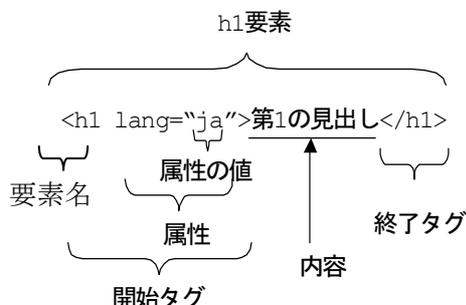
1. 「マイコンピュータ」を開き、メニューバーの[ツール]から[フォルダオプション]を開く。
2. 「表示」タブを選択する。
3. 「登録されている拡張子は表示しない」という項目のチェックをはずす(右図参照)。
4. 「OK」ボタンを押す。これで、拡張子付きのファイル名が表示されるようになる。

※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。



## 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。



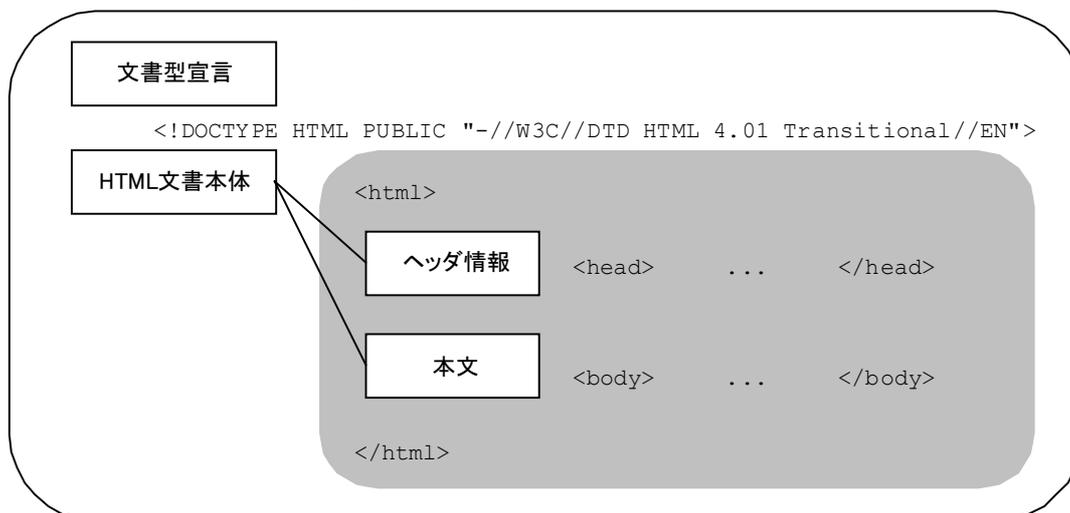
要素名、属性名 (h1 と lang)は必ず半角 (ASCII) で記述すること (大文字, 小文字どちらでもOK。おすすめは全て小文字で入力する方法)。

- 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。

<    &lt; (less than)                    &            &amp; (ampersand)  
>    &gt; (greater than)                "            &quot; (quotation)

- テキストファイルを開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>
  - ◇ 国際経済学部高辻先生の HTML 解説「初めての HTML」:  
<http://www.reitaku-u.ac.jp/tak/tak/ntwk/html.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 の授業で本格的にレッスンするので、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

第 5, 6 回で学習したように、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・および再編集の際は充分注意しよう。

HTML 文書では、< や >, " など、タグ情報は全て ASCII の文字で記述する。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。

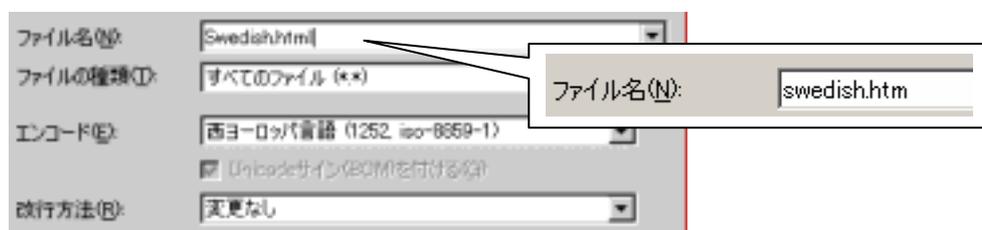
言語・地域で標準的に使われるエンコード方式とコードページ (第 6 回資料より)

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

### 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、エンコード方式として **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「全てのファイル」(または「HTML」)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」という項目がある)。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!** [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、HTML をテキスト文書として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!)
- 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。
- 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ため、1 番目と 3 番目の違いに気づかないことがある!

実習 5: `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。



- 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 <ul>, </li> 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。

アイコンをダブルクリックして WWW ブラウザで表示し、エンコードを正しく指定して正しく表示されるか確認しなさい。

実習 8【実習 6 の復習】: 第 6 回課題として日本語と外国語でそれぞれ作成し、各言語・地域のエンコードで保存した自己紹介のテキストも、同様に HTML 文書にしなさい。

- 日本語は Shift JIS, 外国語はその言語の標準的なエンコードで作成すること。
- ファイル名は以下のように変更すること。(後日提出するのでファイル名を正確につけること。)

言語	元テキストの名前	作成する HTML 文書名	エンコード名
英語	intro_en.txt	学籍番号_en.html	西ヨーロッパ言語
韓国語	intro_ko.txt	学籍番号_ko.html	韓国語
スペイン語	intro_es.txt	学籍番号_es.html	西ヨーロッパ言語
タイ語	intro_th.txt	学籍番号_th.html	タイ語
中国語(台湾)	intro_zh-TW.txt	学籍番号_zh-TW.html	繁体字中国語 (big5)
中国語(中国)	intro_zh-CN.txt	学籍番号_zh-CN.html	簡体字中国語 (GB2312)
ドイツ語	intro_de.txt	学籍番号_de.html	西ヨーロッパ言語
日本語	intro_ja.txt	学籍番号_ja.html	日本語 (シフト JIS)
フランス語	intro_fr.txt	学籍番号_fr.html	西ヨーロッパ言語

学籍番号は半角の数字でつける。学籍番号の後はアンダーバー ( ) が入る。zh-CN, zh-TW の - はハイフンである。

- なお、この 2 つのファイルは以後の実習でも引き続き使うので、次回授業までに内容を充実させておくこと。次回授業後、締め切りを設けて提出してもらう。

2 つの自己紹介は、厳密に同じものでなくともかまわない (日本語ページで外国語ページの内容について簡単に紹介するとよい)。また、内容も評価に加える予定なので、関連する内容のリンク集をページ内に作成しておく、画像を用意する、など各自工夫して内容を充実させよう。

アイコンをダブルクリックして WWW ブラウザで表示する際、また 2 つの言語のページをリンクで行き来すると、エンコードの設定がおかしくなり、自分で正しく指定しなさいなければならない場合があることに気づくだろう。対処方法については次回の授業で取り上げるので楽しみに。

(以上)