
第 11 回：外国語ウェブページの作成 (2)

1. 外国語ウェブページ作成のポイント

1.1. 文字エンコード方式の記述

- HTML 文書がどの文字エンコード方式で作成されているかをヘッダ情報(<head> ... </head> の中)に入れておくことができる。meta 要素は開始タグのみで、内容、終了タグはない(「空要素」と呼ばれる。マークアップの用語については第 10 回資料参照)。以下は日本語 Shift JIS の例 (・ は半角スペースを表す)：

```
<meta http-equiv="Content-Type" content="text/html; charset=Shift_JIS">
```

charset= の直後にあるのが文字エンコード方式の記述である。meta タグの書き方は複雑であるので、とりあえずこのように書くものと覚えておけばよい。日本語以外の言語・地域では、文字エンコード方式の記述が変わる。表記例は以下のとおり：

西ヨーロッパ言語 (英語、スペイン語、ドイツ語、フランス語など)：

```
<meta http-equiv="Content-type" content="text/html; charset=iso-8859-1">
```

タイ語：

```
<meta http-equiv="Content-type" content="text/html; charset=tis-620">
```

韓国語：

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

簡体字中国語：

```
<meta http-equiv="Content-type" content="text/html; charset=gb2312">
```

繁体字中国語：

```
<meta http-equiv="Content-type" content="text/html; charset=big5">
```

Unicode (UTF-8)：

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8">
```

WWW ブラウザの多くは、この情報に基づいて文書の文字エンコード方式を判別するので、この記述があると文字化けが起こりにくくなる。このタグがなければページが正しく表示できない、というわけではないが、文字エンコード方式が分からないときの手がかりになるので、常に記述することをすすめる)¹。

Unicode に対応していない古い WWW ブラウザを利用する人や、日本語などのラテン文字以外のテキストを正しく表示できないブラウザを利用する人も多い。UTF-8 でページを作る場合は、Shift JIS などその言語・地域の標準的なエンコード方式を使ったバージョンを作成しておいたり、Web ページの冒頭で、UTF-8 でエンコードされたテキストである旨を英語 (ASCII のみ) で明記するなど、Unicode テキストを正しく処理できないブラウザのための配慮をすることが望ましい。

実習 1: 上記を参考に English.html にエンコード方式を記述する meta タグを加えなさい。

¹ 第 1 回の授業で閲覧した、自動的にエンコードを判別する多言語ウェブページの例は、この meta タグを使っている。

1.2. 言語名の属性表示

- テキストが何語で書かれているのか、という情報は、文字エンコード方式の表示だけではわかりにくい (ISO-8859-1 のページは、英語？ドイツ語？スペイン語？はたまたフランス語？)。そこで、HTML では、`lang` という属性を使い、より明示的に言語名を記述することができる。言語名属性は、現時点ではまだ利用されているとはいえないが、HTML の規格書でも記述が推奨されている。今後検索ツールなどに活用される可能性があるので、積極的に記述したい (Google など一部実用化が進んでいる)。
- `lang` 属性は、**`lang="言語名"`** という形で開始タグの中に記述する。

例： `<p lang="ja">これは日本語です。</p>`

`lang` 属性の言語名として、言語名表記の国際標準規格 RFC 1766 ² に従った、以下のような言語名を記述するのが一般的である。

- 英語 `en`, `en-UK` (英語, イギリス), `en-US` (英語, アメリカ)
- ドイツ語 `de`
- フランス語 `fr`, `fr-FA` (フランス語, フランス), `fr-CA` (フランス語, カナダ)
- スペイン語 `es`
- タイ語 `th`
- 中国語 `zh`, `zh-CN` (中国の中国語), `zh-TW` (台湾の中国語)
- 韓国語 `ko`
- 日本語 `ja`

日本語は `jp` でなく、`ja` であることに注意。`zh-tw` の `tw` のような言語名の二次属性は国や地域をあらわす (省略可能だが、中国語のように地域によって文字エンコード方式が大きく異なる場合には、つけたほうがよいだろう)。

例： `<html lang="zh-CN">`
`<head>`
`<meta http-equiv="Content-type" content="text/html; charset=gb2312">`
`<title lang="en">Introducing Shoju CHIBA in Chinese</title>`
`</head>`
`<body>`
`<p lang="en">This page is written in Simplified Chinese (GB2312).</p>`
`<h1>我是千叶, 你好吗!</h1>`
`...`

² RFC は Request For Comments の略で、インターネットの技術標準を提案する団体 IETF (Internet Engineering Task Force) によって策定、公開されている。インターネットにかかわる技術的な実装のほとんどは RFC に準拠しておこなわれており、RTF は事実上インターネットの技術仕様書としての意味をもっている。このような標準化推進団体には、HTML を策定している W3C (World Wide Web Consortium) や Unicode の規格策定と普及のための活動をおこなっている Unicode Consortium などがある。

RFC1766 の正式名称は "Tags for the Identification of Languages" で IETF のホームページ (<http://www.ietf.org/rfc/rfc1766.txt>) で読むことができる (ただし、内容はテクニカルである)。表記は言語の表記 (前半部) を ISO 639 ("Code for the representation of names of languages", <http://www.ics.uci.edu/pub/ietf/http/related/iso639.txt>) による 2 字の表記に従い、国・地域名の表記 (後半部) は ISO 3166 ("Codes for the representation of names of countries", <http://www.ics.uci.edu/pub/ietf/http/related/iso3166.txt>) に従って記述するのが現在のところ一般的である。(1988 年に策定された ISO639 は約 130 足らずの言語しか規定しておらず、その後改訂され、400 あまりの言語を含む 3 文字のコードが 1998 年に ISO639-2 として規定された。)

```
</body>
</html>
```

実習 2：§ 1.2. を参考に，English.html の html タグ，title タグに lang 属性を使って言語名 (en-UK) を記述しなさい。

1.3. 文字参照・実体参照を使った文字表記

- HTML 文書では，普段あまり使わない文字など，ある特定の文字を「文字参照」というしくみを使って文書中に呼び出すことができる。文字参照には 2 種類あり，使用したい文字の文字コード (Unicode で表記) を 10 進数または 16 進数で表記する「数値文字参照」と，文字をあらかじめ決められた略称で表現する「文字実体参照」がある。前者は Unicode に存在している文字であれば原則として何でも表記できる。一方後者は西ヨーロッパ言語 ISO-8859-1 に収録された文字について利用できる (**補足資料参照**，前回の配布資料 § 2.2. で挙げた <, >, &, " の 4 つの記号の置き換え規則も文字実体参照を使っている)。

数値文字参照を使うと，編集時の言語の文字エンコード方式で表現できない文字を表示させることができる。たとえば，日本語の Web ページ中でも，日本語 (Shift JIS) にない文字を表示させることができる。たとえば，ユーロ記号 (€) の Unicode での文字コードは 16 進法で「20AC」であるが，`€` という表記 (10 進法なら `€`) を使えば Shift JIS など，ローカルな文字エンコード方式で作成されたテキストの中でユーロ記号を挿入することができる (Unicode の文字の検索方法やコードの調べ方についての詳細は，第 6 回資料を，10 進数と 16 進数の変換は第 7 回資料を参照)。

参照文字を使った HTML の例：

文字実体参照：

```
<p>フィンランド語で「お母さん」は <i>&auml;iti</i> といいます。</p>
```

数値文字参照(16 進数)：

```
<p>フィンランド語で「お母さん」は <i>&#x00E4;iti</i> といいます。</p>
```

数値文字参照(10 進数)：

```
<p>フィンランド語で「お母さん」は <i>&#228;iti</i> といいます。</p>
```

ブラウザでの表示例：

フィンランド語で「お母さん」は äiti といいます。

ブラウザによっては，参照を使って呼び出した文字が正しく表示されないことがある (例えば，16 進数表記の数値文字参照は大学の Netscape (バージョン 4.78) ではうまくいかず，前回挙げた 4 つの記号，<, >, &, " の文字参照しか正しく表示できない)。

実習 3：§ 1.3. を参考に，数値文字参照を使い English.html にユーロ記号を入れなさい。

```
<li>Price: UK £6.99, EUR &#x20AC;12.23</li>
```

提出課題 (Web ページ作成)

- 前回の実習 7 で作成した 2 言語の自己紹介の HTML 文書に，§ 1.1. を参考にエンコード方式を記述する meta タグを加えなさい。
- 同様に，§ 1.2. を参考に，html タグ，title タグに lang 属性を使って言語名を記述しなさい。

- 他のページへのリンクを作成したり，画像を使ったりして，自己紹介のウェブページのコンテンツを充実させなさい。
- 作成した自己紹介のウェブページを Kadai サーバの [schiba] → [2005fl] → [html] フォルダの [class1] (1 クラス (千葉))，[class2] (2 クラス (西田)) というフォルダにネットワーク経由で提出しなさい。画像や追加で作成した HTML 文書など，一緒に提出する他のファイルがある場合には，学籍番号の名前をつけたフォルダを作成してファイルをその中にまとめて入れ，フォルダごと提出するとよい。1 クラスの提出締め切りは 6 月 29 日 (火) 中とする (2 クラスは教員の指示に従うこと)。遅れると提出できないので注意。

(以上)

補足資料 (§ 1.3. 参照) :

大 藤 幹 (2005) 『詳解 HTML & XHTML & CSS 辞典』改訂版 . 秀和システム .

PDF 資料では省略

大藤幹 (2005) 『詳解 HTML&XHTML&CSS 辞典』改訂版 . 秀和システム .