Babel 2012 on the Web

Daniel Glazman
21-sep-2011

W3C Workshop
A Local Focus for the Multilingual Web
1. Where we come from...
« Is it really important to support Boustrophedon or Mongolian in CSS? »

CSS Working Group, June 1998

« Since many countries use characters which are not a part of ASCII, the default character-set for modern browsers is ISO-8859-1 »

W3Schools, now...
Top Ten Languages in the Internet
2010 - in millions of users

- English: 536.6
- Chinese: 444.9
- Spanish: 153.3
- Japanese: 99.1
- Portuguese: 82.5
- German: 75.2
- Arabic: 65.4
- French: 59.8
- Russian: 59.7
- Korean: 39.4
- All the rest: 350.6

Estimated Internet users are 1,955,514,818 on June 30, 2010
Copyright © 2000 - 2010, Minhwaits Marketing Group
<table>
<thead>
<tr>
<th>US-ascii</th>
<th>UTF-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>uuencode</td>
<td>MIME</td>
</tr>
<tr>
<td>7bits</td>
<td>8bits</td>
</tr>
<tr>
<td>URLs</td>
<td>IRIs</td>
</tr>
<tr>
<td>-</td>
<td>Accept-Language</td>
</tr>
<tr>
<td>HTML+</td>
<td>HTML5</td>
</tr>
<tr>
<td>-</td>
<td>CSS 3</td>
</tr>
<tr>
<td>-</td>
<td>xml:lang</td>
</tr>
</tbody>
</table>
2. on the radar today...
Richard Ishida's business card...
A reminder...

language ≠ script ≠ charset
HTML 5: charset

• ...Authors are encouraged to use UTF-8. Conformance checkers may advise authors against using legacy encodings ...

• ...Authoring tools should default to using UTF-8 for newly-created documents ...
HTML5: language

• still `lang` and `xml:lang`...
• the lang IDL reflects only `lang`...
• Authoring Tools rarely set the language or even offer UI for it
HTML5: links

- hyperlinks in HTML are still mono-valued
- the hreflang attribute can target only one language
HTML5: direction

- the `dir` attribute cannot reflect vertical directions (yet)
- elements with different inner and outer directions still a problem
- `:ltr` and `:rtl` pseudo-classes...
- `bdi` and `bdo` elements poorly implemented
HTML5: forms

- bad localizations of dates and calendars
- date/time issues with time zones
- what's a name...
JavaScript

- poor localizability...
- user interaction entirely based on UA's language and direction
- Node.js spreading!
DOM: charset

• DOM uses UTF-16 strings...
PHP

- PHP 5 said to have a lot of issues with UTF-8
- PHP 6 should fix that (hopefully...)
- Poor built-in localizability
CSS3: Writing Modes

10月27日、マイクロソフトの定番ビジネス・スイート「マイクロソフト・オフィス」の最新版となる「オフィス2011（Office for Mac 2011）」がついに発売されました。2009年8月、マイクロソフトは新しいMac版オフィスについてのロードマップを示しました。当時、2010年のソースに基づく「メイクオーバー」に発売する予定であることが、新たに統合されたOffice 2010の「アウトルック」を搭載することなどが発表されましたが、詳細はあまり明らかにされず、Macユーザは長い間その詳細の発表を心待ちにしていました。

Figure 2. Mongolian-based writing mode
CSS3: Text

- text transformations (issues in Greek)
- hyphenation
- emphasis marks (mostly for East Asian text)
CSS3: Columns
CSS3: Lists

- list-style-type property extended to dozens of values
- ability to define your own if missing
CSS3: Box Model

- start/end instead of left/right
- margin-{start|end} not here yet
CSS3: Fonts

- Language-specific display
  
  Señora ➤ Señora

- Control of glyph substitute and positioning in East Asian text
  
  麗町 ➤ 麗町
CSS3: Ruby
EPUB3

- OpenType & WOFF
- CSS 2.1 + parts of CSS 3
  - hyphenation, emphasis, word breaking
  - Writing Modes
  - Multi-column Layout
  - Ruby
3. what we can expect...
This happened to my friend:

English should be the national language. These immigrants should have to learn English when they come here.

Yeah.

When you go to live somewhere, you learn the language they speak there.

English is the language of the land.

Excuse me, but osio Sarah dawado.

What the hell was that?

Cherokee.

CC BY-NC 2.5 [XKCD]
• HTML5+CSS3 as the pivot format for new Wysiwyg editors with good i18n
• massive adoption of EPUB3 in Asia
• Boustrophedon 😐:-)
Thank you