

Babel 2012 on the Web

Daniel Glazman

21-sep-2011

W3C Workshop
A Local Focus for the Multilingual Web



I. Where we come from...

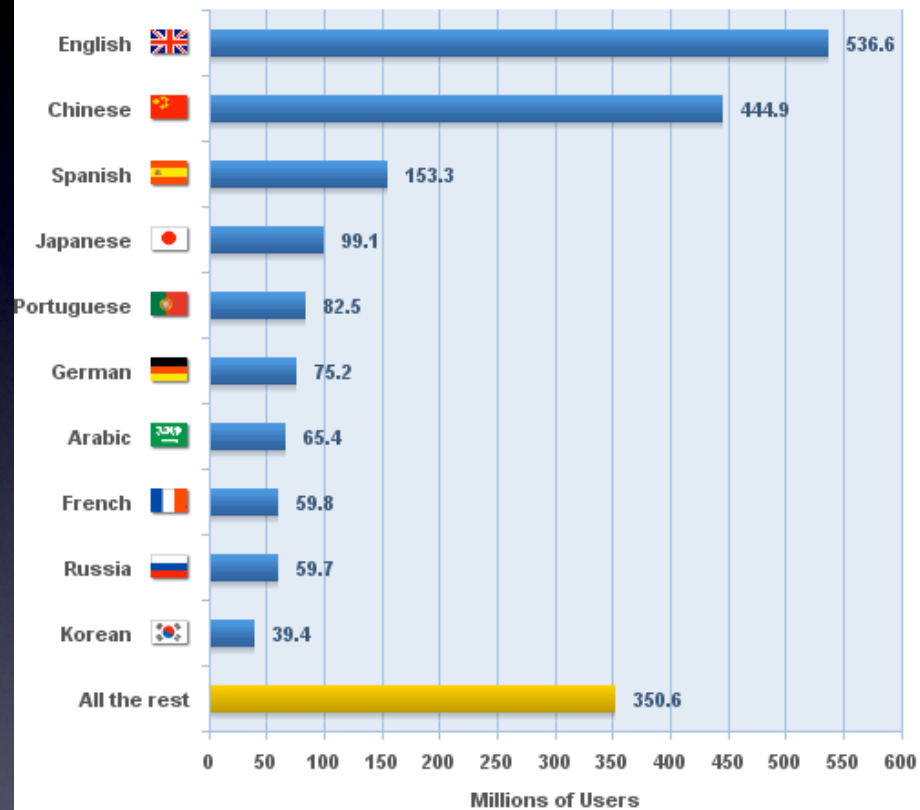
« Is it really important to
support Boustrophedon or
Mongolian in CSS? »

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



Source: Internet World Stats - www.internetworldstats.com/stats7.htm
Estimated Internet users are 1,966,514,816 on June 30, 2010
Copyright © 2000 - 2010, Miniwatts Marketing Group

US-ascii	UTF-8
uuencode	MIME
7bits	8bits
URLs	IRIs
-	Accept-Language
HTML+	HTML5
-	CSS 3
-	xml:lang

2. on the radar today...

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

```
<ul>
  <li>User <bdi>jcranmer</bdi>: 12 posts.
  <li>User <bdi>hober</bdi>: 5 posts.
  <li>User <bdi>إيان</bdi>: 3 posts.
</ul>
```

- User jcranmer: 12 posts.
- User hober: 5 posts.
- User 3 إيان: 3 posts.

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年のホリデーシーズンに発売する予定であることや、新たに統合メールソフトの「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たわけです。

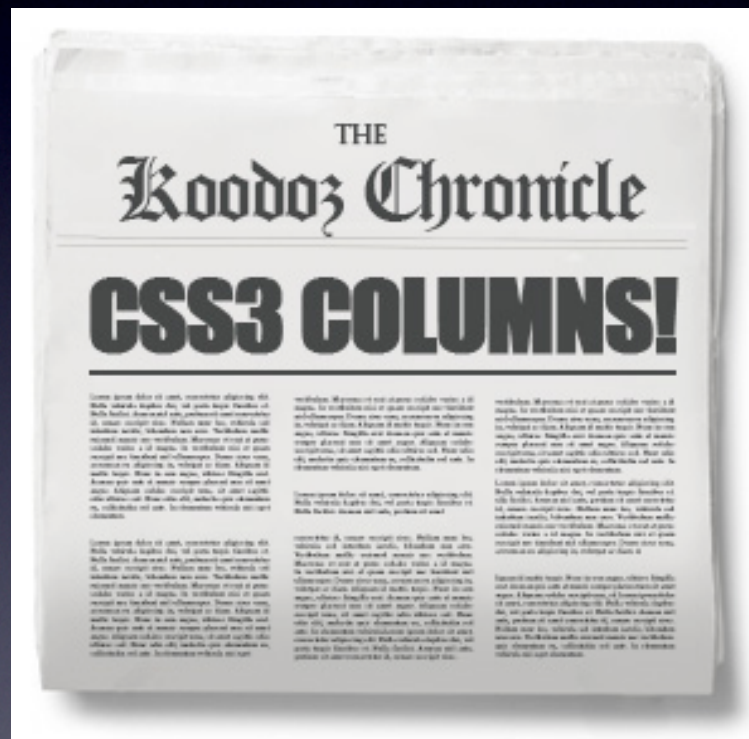


Figure 2. Mongolian-based writing mode

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

麴町 ▶ 麴町

CSS3: Ruby

とう きょう
東 京

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:

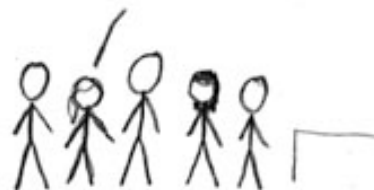
English should be the national language. These immigrants should have to learn English when they come here.



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



Excuse me, but osio Sarah dawado.



What the hell was that?

Cherokee.)



- HTML5+CSS3 as the pivot format for new Wysiwyg editors with good i18n
- massive adoption of EPUB3 in Asia
- Boustrophedon ? :-)

Thank you

ありがとう

Bbl

merc

σας ευχαριστούμε

obrigado

gracias

grazie

danke

dank u

당신을 감사하십시오