HTML5 i18n: A report from the front line

Richard Ishida
W3C Internationalization Activity

Jirka Kosek
University of Economics, Prague
Outline

W3C Internationalization Activity
Bidirectional text
Ruby
Translate flag
About the Consortium


Mission:
Lead the technical evolution of the Web and ensure its interoperability.
W3C technology stack
Internationalization (i18n) Core Working Group

• Help W3C Working Groups understand issues and build in requirements relating to worldwide support for Web technologies

• Liaise with other standards organizations to develop support for the international Web (Unicode, IETF, etc)

• Help users of Web technology understand what's available to them and how to use it by developing best practices and other resources
MultilingualWeb-LT Working Group

- Define meta-data for web content (mainly HTML5) and "deep Web" content, for example a CMS or XML files from which HTML pages are generated, that facilitates its interaction with multilingual technologies and localization processes.

- Demonstrate interoperable implementations.
Internationalization Interest Group

• www-international@w3.org
• public-i18n-bidi@w3.org
• public-i18n-cjk@w3.org
• public-i18n-indic@w3.org
• public-i18n-its@w3.org
• public-iri@w3.org
Bidi
<table>
<thead>
<tr>
<th>TOP RATED RESTAURANTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aroma - 3 reviews</td>
</tr>
<tr>
<td>피자 스포일러</td>
</tr>
<tr>
<td>Roma - 3 reviews</td>
</tr>
<tr>
<td>피자 스포일러</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TOP RATED RESTAURANTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aroma - 3 reviews</td>
</tr>
<tr>
<td>피자 스포일러</td>
</tr>
<tr>
<td>5 - 5 reviews</td>
</tr>
<tr>
<td>Roma - 3 reviews</td>
</tr>
<tr>
<td>피자 스포일러</td>
</tr>
</tbody>
</table>
<p><bdi>Aroma</bdi> - 3 reviews</p>
<p><bdi>PURPLE PIZZA</bdi> - 5 reviews</p>
Ruby
Ruby
What is it?
Requirements for Japanese Text Layout

W3C Working Group Note 4 June 2009

This version:
http://www.w3.org/TR/2009/NOTE-jitreq-20090604/
Latest version:
http://www.w3.org/TR/jitreq/
Previous version:
http://www.w3.org/TR/2008/WD-jitreq-20081015/
Editors:
Yasuhiro Anan (Navara Dentsu), Microsoft
Hiroyuki Chiba (千葉 弘幸), Invited Expert
Junshiro Edamoto (杉本 典隆), Invited Expert
Richard Ishida, W3C
Kenichiro Ishino (石野 一郎), Antenna House
Tatsuo Kobayashi (小林 眞), JustSystems
Toshi Kobayashi (小林 哲), Invited Expert
Keizou Onozawa (小野澤 敖), Invited Expert
Felix Sasaki, University of Applied Sciences Potsdam

Please refer to the errata for this document.
A Japanese version of this document is also available. See also translation.
English version of this document is the authoritative version.

Copyright © 2009-2009 W3C (WHAT WORKS). All Rights Reserved. W3C liability, trademarks, and document use rules apply.

Abstract

This document describes requirements for general Japanese layout realized with technologies like CSS, SVG and XSL-FO. The document is mainly based on a framework for Japanese layout, JIS X 4051, however, it also addresses areas which are not covered by JIS X 4051.
Ruby
HTML5 innovations

XHTML
<ruby>
<rb>日本</rb>
<rt>にっぽん</rt>
</ruby>
<ruby>
<rb>電</rb>
<rt>でん</rt>
</ruby>
<ruby>
<rb>気</rb>
<rt>き</rt>
</ruby>

HTML5
<ruby>
日本
<rt>にっぽん</rt>
</ruby>
<ruby>
電
<rt>でん</rt>
</ruby>
<ruby>
気
<rt>き</rt>
</ruby>
Use case
Ruby base styling

B
<rt>a</rt>
B
<rt>a</rt>
<ruby>
Use case
Ruby base styling

<ruby>
<span>B</span><span>a</span></ruby>
<ruby>
<rt>a</rt>B
<rt>a</rt>
<ruby>
Use case
Ruby base styling

• Use of span works.
• If used a lot, less obvious that this is better than rb.
• But allows styling within a base unit of multiple characters.
• Only works where span is present, so mass conversion for accessibility is not possible – but then neither does rb help if optional.

• Possible alternative, implicit rb element.
今日は会議の初日です。

今日は会議（かいぎ）の初日です。

今日は会議（かいぎ）の初日です。

B
<ruby>
B
<rp>(</rp>
<rt>a</rt>
<rp>)</rp>
B
<rp>(</rp>
<rt>a</rt>
<rp>)</rp>
<ruby>
Use case
Fallback
今日は会議の初日です。
(今日は会議の初日です。)
(会議の初日です。)
(かいぎ) (かいぎ)
(かいぎ) (かいぎ)
Use case

Fallback

• Should we go to a column-based, rather than row-based approach? This would involve introducing rb.
• Is there some other way of producing a fallback?
• Can we reach a situation quickly where fallbacks are not needed?
• some parts of texts should not be translated
• nowadays translation is usually automated thus there is a need for machine readable information about translatability

<p>Click the Resume button on the Status Display or the CONTINUE button on the printer panel.</p>

CONTINUE is a label on a hardware button and should not be translated as labels on hardware buttons are not going to be localized.
Translate flag

HTML5 solution

- translate="no" attribute can be added to any HTML element to indicate that content of element should not be translated

<p>Click the Resume button on the Status Display or the CONTINUE button on the printer panel.</p>

<p>Click the Resume button on the Status Display or the <span translate="no">CONTINUE</span> button on the printer panel.</p>
• translate attribute was recently added to HTML5 by effort started at the MLW Pisa workshop in 2011
• it is available in many other XML based formats thanks to ITS (Internationalization Tag Set)
• translate attribute can have yes and no values which can be overridden
• already supported by on-line translation services from Microsoft and Google
• already supported by content formats like DITA or DocBook
Getting involved…
The Web needs your help

d this is your Web – not the W3C's

de the Web is about people, not technology

we need You to make the Web worldwide

g get involved

Thank you

http://www.w3.org/International/