The Multilingual Web: Latest developments at the W3C/IETF

Richard Ishida
W3C Internationalization Activity

Outline

About the W3C

Standards support for the multilingual Web

The changing social context

Best practices for the multilingual Web

Getting involved

Outline

Why is it that in 2011, it is still difficult for users and developers around the world to use the Web for their own language and culture?

Which issues are more or less solved on the web (and how)?

What are we doing to address the remaining problems, and how can you influence the outcomes?

About the W3C Internationalization Activity

About the Consortium

1994: World Wide Web Consortium created and still led by:
Sir Tim Berners-Lee,
inventor of the WWW.

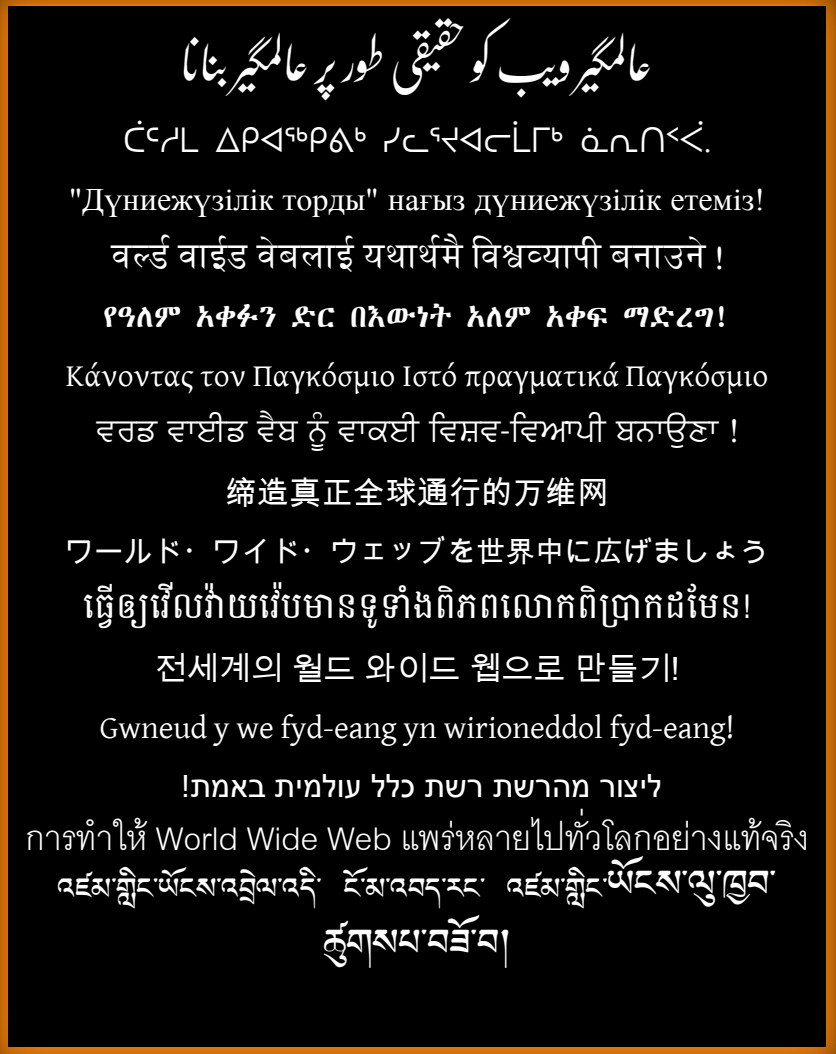
Mission:
Lead the
technical evolution of the Web
and ensure its
interoperability

Keywords:
consensus and vendor neutrality

About the Consortium

Internationalization Activity

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



Standards support for the multilingual Web

Standards support

Unicode

جعل شبكة الويب العالميّة عالميّة حقًّا!

وب جهانی را به درستی جهانی سازیم!

حقیقی طور پر عالمگیر بنانا

Համաշխարհային ցանցն իրոք համաշխարհային դարձնելը

ᑖᑦᓱᒪ ᐃᑭᐊᖅᑭᕕᒃ ᓯᓚᕐᔪᐊᓕᒫᒥ

"Дүниежүзілік торды" нағыз дүниежүзілік етемез!

वर्ल्ड वाईड वेबलाई यथार्थमै विश्वव्यापी बनाउन !

የዓለም አቀፉን ድር በእውነት አለም አቀፍ ማድረግ!

Κάνοντας τον Παγκόσμιο Ιστό πραγματικά Παγκόσμιο

ਵਰਡ ਵਾਈਡ ਵੈਬ ਨੂੰ ਵਾਕਈ ਵਿਸ਼ਵ-ਵਿਆਪੀ ਬਨਾਉਣਾ !

缔造真正全球通行的万维网

ליצור מהרשת רשת כלל עולמית באמת!

ˈmeɪkɪŋ ðə wɜːld waɪd wɛb ˈtruːlɪ ˈwɜːldˈwaɪd

ワールド・ワイド・ウェッブを世界中に広げましょう

ធ្វើឲ្យវើលវ៉ាយវ៉េបមានទូទាំងពិភពលោកពិប្រាកដមែន!

전세계의 월드 와이드 웹으로 만들기!

Gwneud y we fyd-eang yn wirioneddol fyd-eang!

การทำให้ World Wide Web แพร่หลายไปทั่วโลกอย่างแท้จริง

འཛམ་གླིང་ཡོངས་འབྲེལ་འདི་ ངོ་མ་འབད་རང་ འཛམ་གླིང་ཡོངས་ལུ་ཁྱབ་ཚུགས་པ་བཟོ་བ།

"The Path W3C follows to making text on the Web truly global is Unicode."

Tim Berners-Lee

Standards support

Unicode

جعل شبكة الويب العالميّة عالميّة حقًّا!

وب جهانی را به درستی جهانی سازیم!

عالمگیر ویب کو حقیقی طور پر عالمگیر بنانا

Համաշխարհային ցանցն իրոք համաշխարհային դարձն

ᑖᑦᓱᒪ ᐃᑭᐊᖅᑭᕕᒃ ᓯᓚᕐᔪᐊᓕᒫᒥᒃ ᓈᓇᑎᐸᐹ.

"Дүниежүзілік торды" нағыз дүниежүзілік етеміз!

वर्ल्ड वाईड वेबलाई यथार्थमै विश्वव्यापी बनाउने !

የዓለም አቀፉን ድር በእውነት አለም አቀፍ ማድረግ!

Κάνοντας τον Παγκόσμιο Ιστό πραγματικά Παγκόσμιο

ਵਰਡ ਵਾਈਡ ਵੈਬ ਨੂੰ ਵਾਕਈ ਵਿਸ਼ਵ-ਵਿਆਪੀ ਬਨਾਉਣਾ !

缔造真正全球通行的万维网

ליצור מהרשת רשת כלל עולמית באמת!

ˈmeɪkɪŋ ðə wɜːld waɪd wɛb ˈtruːlɪ ˈwɜːldˈwaɪd

ワールド・ワイド・ウェッブを世界中に広げましょう

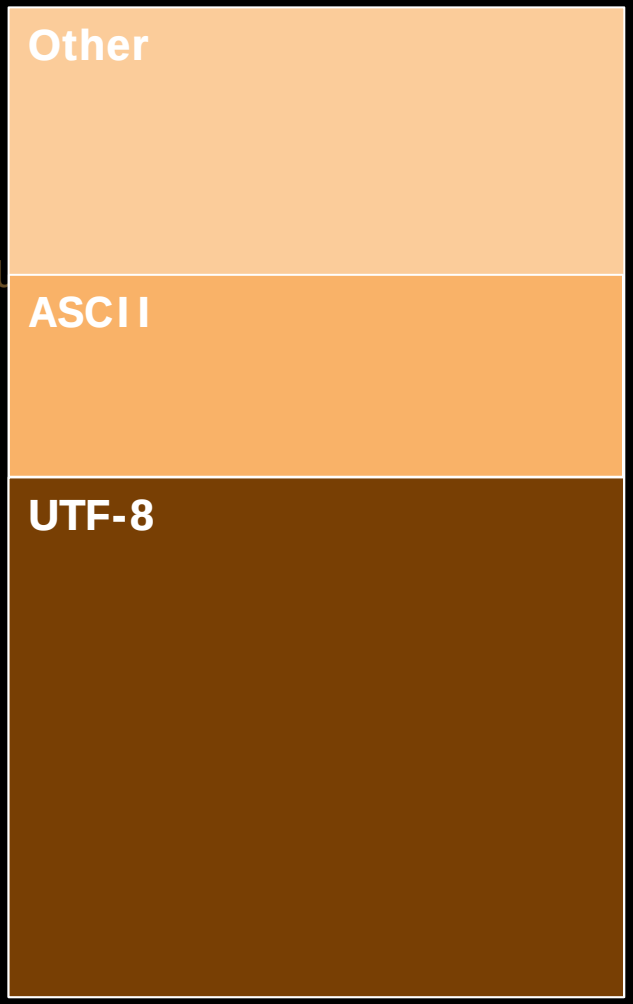
ធ្វើឲ្យវើលវ៉ាយវ៉េបមានទូទាំងពិភពលោកពិប្រាកដមែន!

전세계의 월드 와이드 웹으로 만들기!

Gwneud y we fyd-eang yn wirioneddol fyd-eang!

การทำให้ World Wide Web แพร่หลายไปทั่วโลกอย่างแท้จริง

འཛམ་གླིང་ཡོངས་འབྲེལ་འདི་ ངོ་མ་འབད་རང་ འཛམ་གླིང་ཡོངས་ལུ་ཁྱབ་ཚུགས་པ་བཟོ་བ།



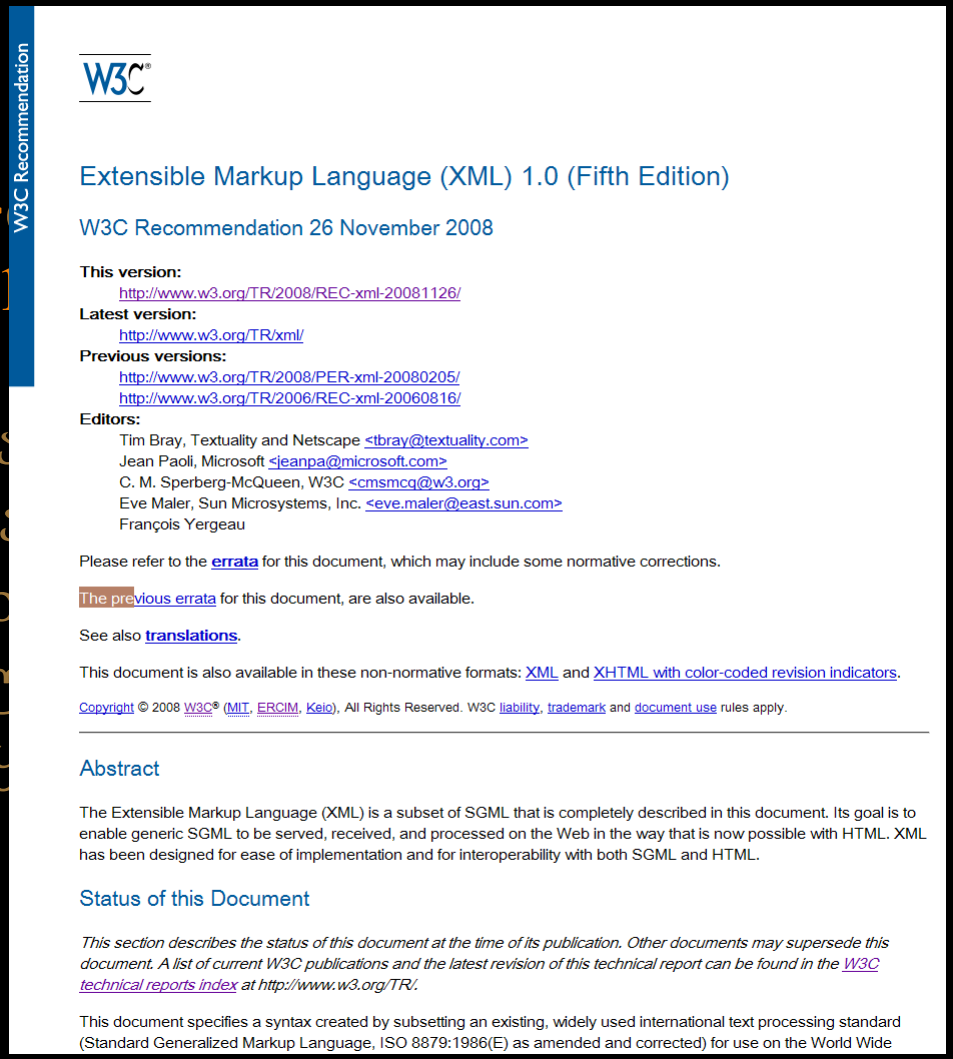
Unicode on the Web

Standards support

Unicode

<h2>
ფოტოსურათი</h1

<p><a href="/wiki/ჭიათ
class="mw-redirect">ჭი
ხოჭოების ოჯახს ეკუ
მომრგვალო ან ოვალუ
ფონზე შავი ლაქები აყ
...



W3C Recommendation

W3C

Extensible Markup Language (XML) 1.0 (Fifth Edition)

W3C Recommendation 26 November 2008

This version:
http://www.w3.org/TR/2008/REC-xml-20081126/
Latest version:
http://www.w3.org/TR/xml/
Previous versions:
http://www.w3.org/TR/2008/PER-xml-20080205/
http://www.w3.org/TR/2006/REC-xml-20060816/
Editors:
Tim Bray, Textuality and Netscape <tbray@textuality.com>
Jean Paoli, Microsoft <jeanpa@microsoft.com>
C. M. Sperberg-McQueen, W3C <cmsmcq@w3.org>
Eve Maler, Sun Microsystems, Inc. <eve.maler@east.sun.com>
François Yergeau

Please refer to the **errata** for this document, which may include some normative corrections.

The previous errata for this document, are also available.

See also **translations**.

This document is also available in these non-normative formats: XML and XHTML with color-coded revision indicators.

Copyright © 2008 W3C® (MIT, ERCIM, Keio), All Rights Reserved. W3C liability, trademark and document use rules apply.

Abstract

The Extensible Markup Language (XML) is a subset of SGML that is completely described in this document. Its goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML.

Status of this Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of current W3C publications and the latest revision of this technical report can be found in the W3C technical reports index at http://www.w3.org/TR/.

This document specifies a syntax created by subsetting an existing, widely used international text processing standard (Standard Generalized Markup Language, ISO 8879:1986(E) as amended and corrected) for use on the World Wide

Standards support

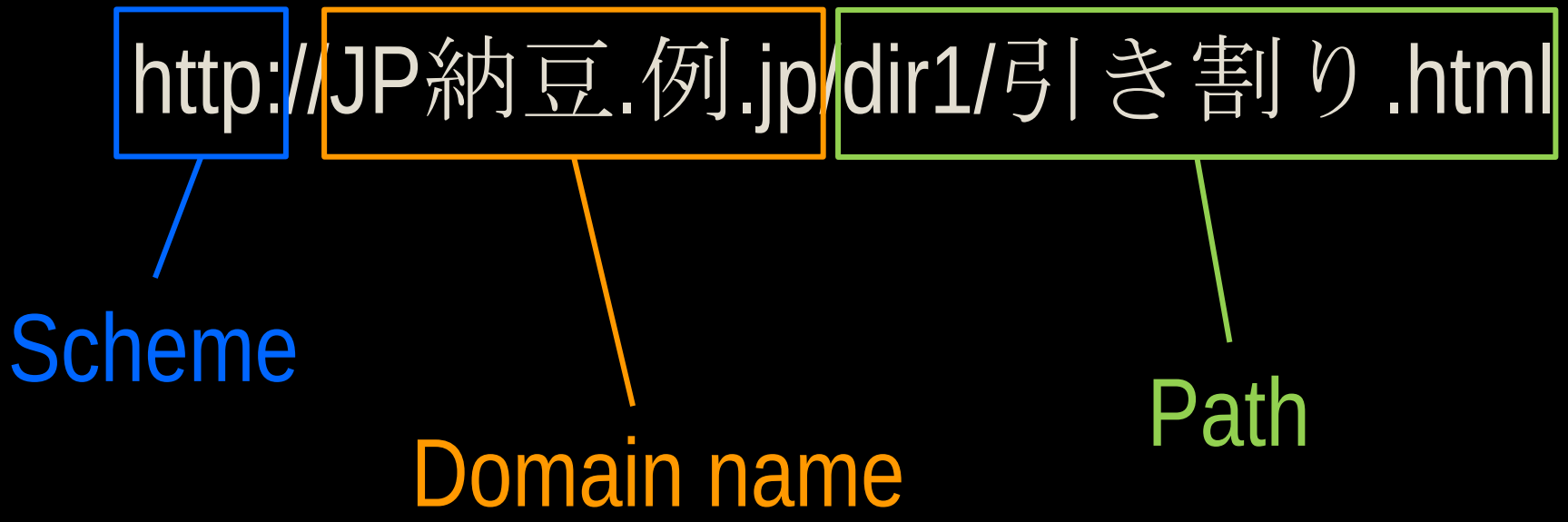
Unicode normalization

NFD

NFC

Standards support

Web resource identifiers



IDN

xn--jp-cd2fp15c.xn--fsq.jp

Standards support

Top level domain names

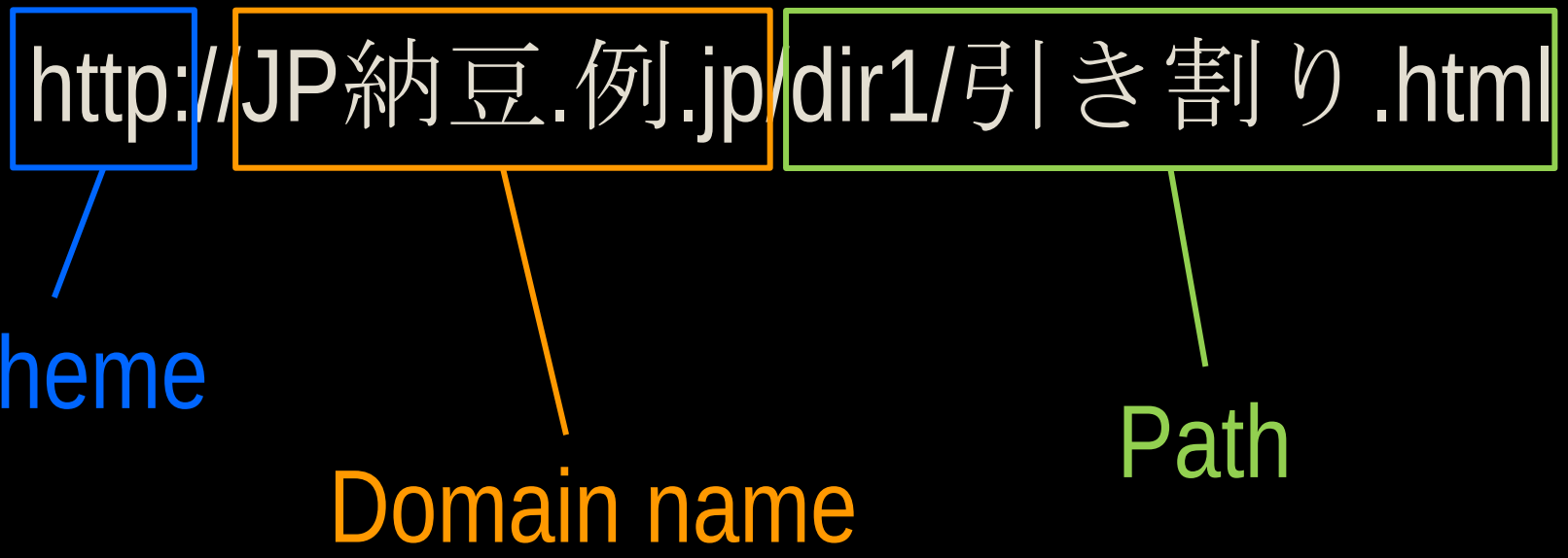
السعودية	Al-Saudiah
امارات	Emarat
مصر	Misr

http://وزارة-الأتصالات.مصر

IDN

Standards support

Web resource identifiers



IRI

/dir1/%E5%BC%95%E3%81%8D%E5%89%B2%E3%82%8A.html

Standards support

Language tags

Before (RFC 3066)

language – region

- ISO 639 language codes
- ISO 3166 country codes

en

en-GB

en-scouse

Standards support

Language tags: BCP 47

Now BCP 47

language script region variant extension private_use

(extlang)

- nearly 8,000 subtags available
- subtags available only from new IANA registry (based on ISO and UN codes)
- only language subtag required

hi

az-Cyrl

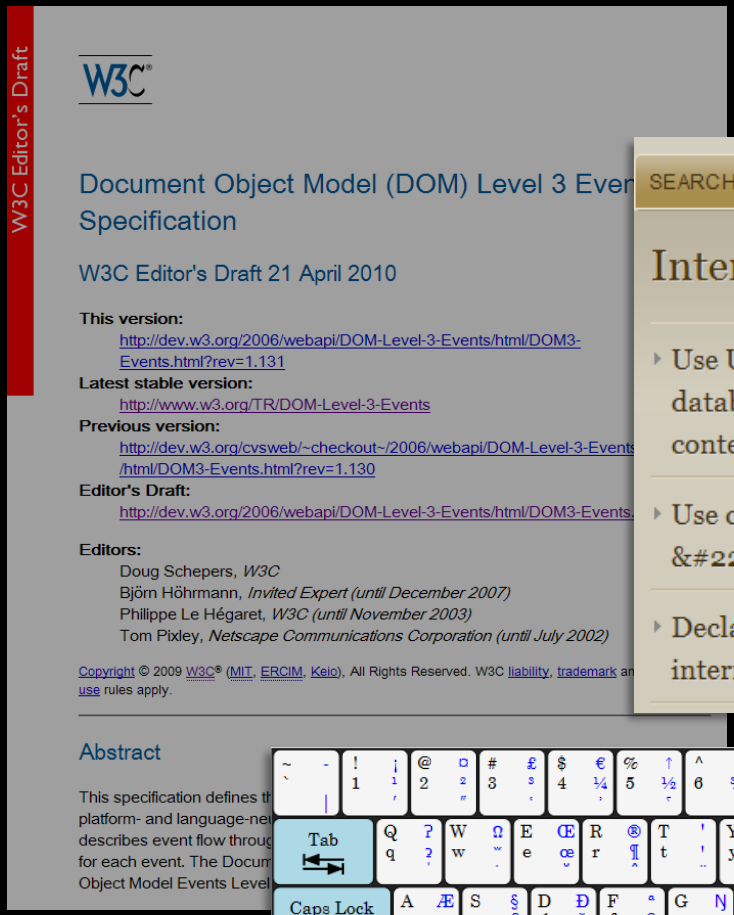
zh-Hans

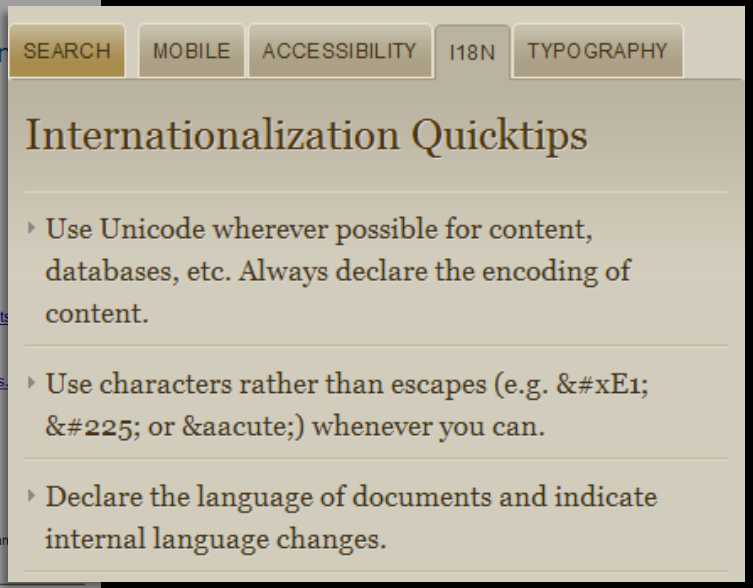
es-419

sl-IT-rozaj-njiva-1994

Standards support

Key Events





Developing requirements

Speech Synthesis Markup Language

W3C Proposed Recommendation

W3C

Speech Synthesis Markup Lan

W3C Proposed Recommendation 23 F

This version:
http://www.w3.org/TR/2010/PR-speech-synthesis11-20100223/
Latest version:
http://www.w3.org/TR/speech-synthesis11/
Previous version:
http://www.w3.org/TR/2009/CR-speech-synthesis11-20090827/

Editors:
Daniel C. Burnett, Voxeo (formerly of Vocalocity and Nuance)
双志伟 (Zhi Wei Shuang), IBM
Authors:
Paolo Baggia, Loquendo
Paul Bagshaw, France Telecom
Michael Bodell, Microsoft
胡方 (Hu Fang), Invited Expert (until 20 October 2009)
黄德智 (De Zhi Huang), France Telecom
黄力行 (Lixing Huang), Chinese Academy of Sciences
康永国 (Yongguo Kang), Panasonic Corporation (until 5 December 2007)
楼晓雁 (Lou Xiaoyan), Toshiba
Scott McGlashan, HP
蒙美玲 (Helen Meng), Chinese University of Hong Kong (until 29 July 2009)
陶建华 (Jianhua Tao), Chinese Academy of Sciences
王霞 (Wang Xia), Nokia (until 30 October 2006)
夏海荣 (Xia Hairong), Panasonic Corporation (until 2 August 2006)
吴志勇 (Zhiyong Wu), Chinese University of Hong Kong (until 29 July 2009)
严峻 (Yan Jun), iFLYTEK

Copyright © 2010 W3C® (MIT, ERCIM, Keio), All Rights Reserved. W3C liability, trademark and document use rules a

Abstract

The Voice Browser Working Group has sought to develop standards to enable access to the Web using spoken interaction. The Speech Synthesis Markup Language Specification is one of these standards and is designed to provide a rich, XML-based markup language for assisting the generation of synthetic speech in Web and other applications. The essential role of the markup language is to provide authors of synthesizable content a standard way to control aspects of speech such as pronunciation, volume, pitch, rate, etc. across different synthesis-capable platforms.

這一晚會如常舉行

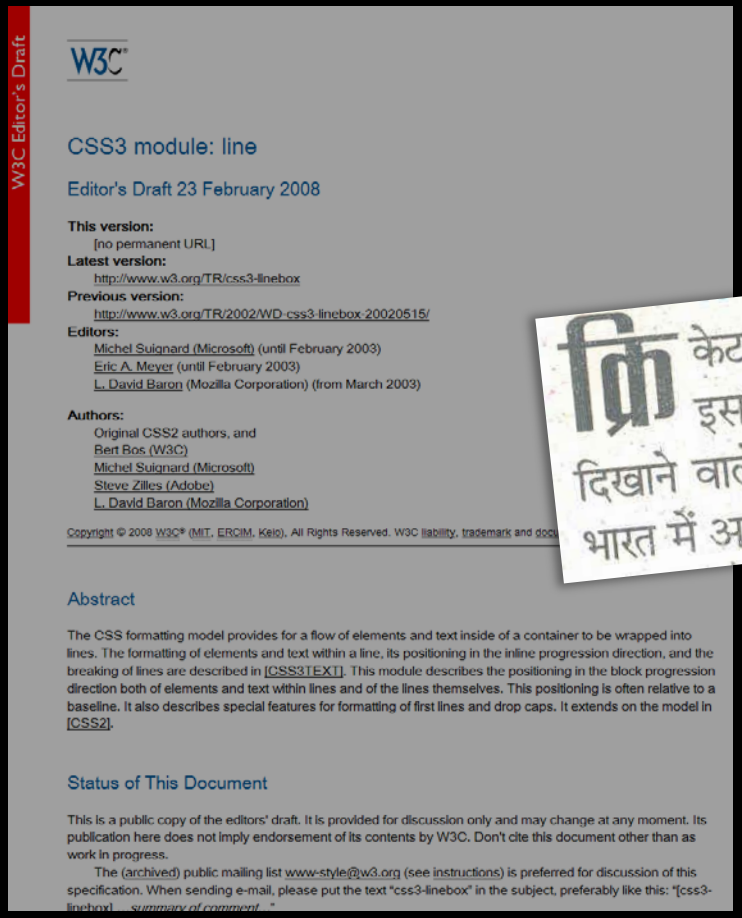
這一|晚會|如常|舉行

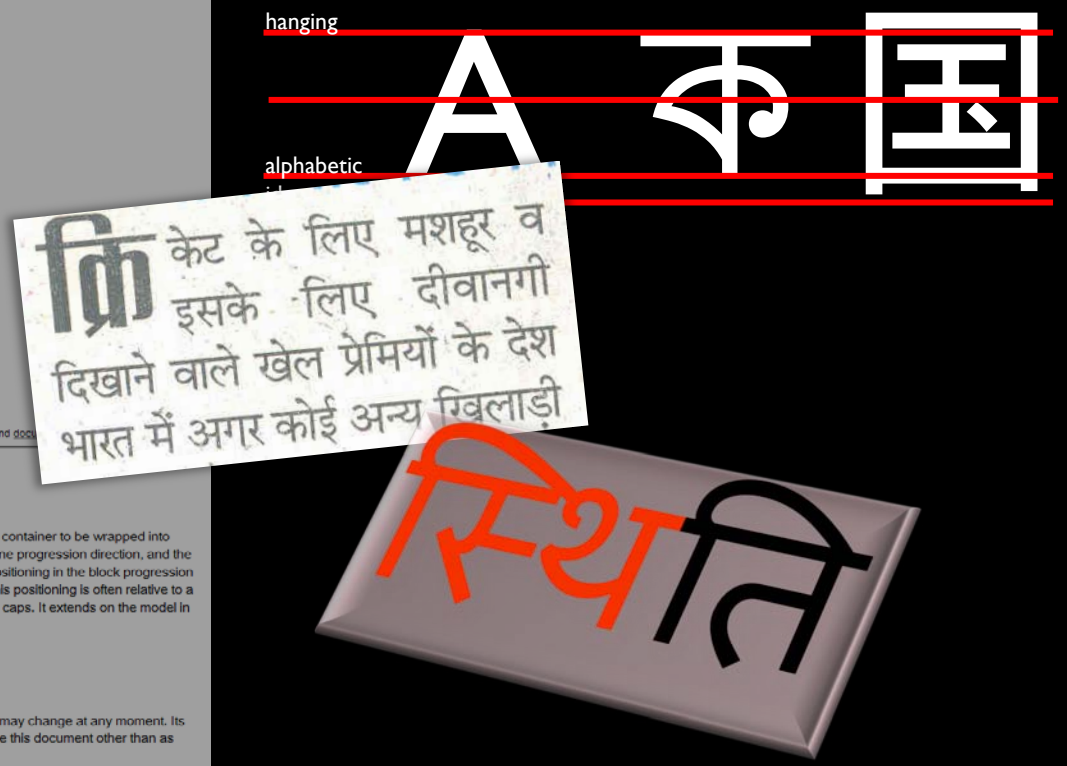
這一|晚會|如|常|舉行

這一晚|會|如常|舉行

Standards support

CSS3





Implementers of user agents need to be prodded by the public to support the developing marketplace !

Standards support

Hyphenation

Zusätzlich erleichtert PLS die Eingrenzung von Anwendungen, indem es Aussprachebelange von anderen Teilen der Anwendung abtrennt.

Zusätzlich er-
leichtert PLS die
Eingrenzung von
Anwendungen, in-
dem es Aussprac-
hebelange von an-
deren Teilen der
Anwendung ab-
trennt.

```
* { hyphens: auto; }
```

Standards support

OpenType feature support by language

Београд, Април 1944

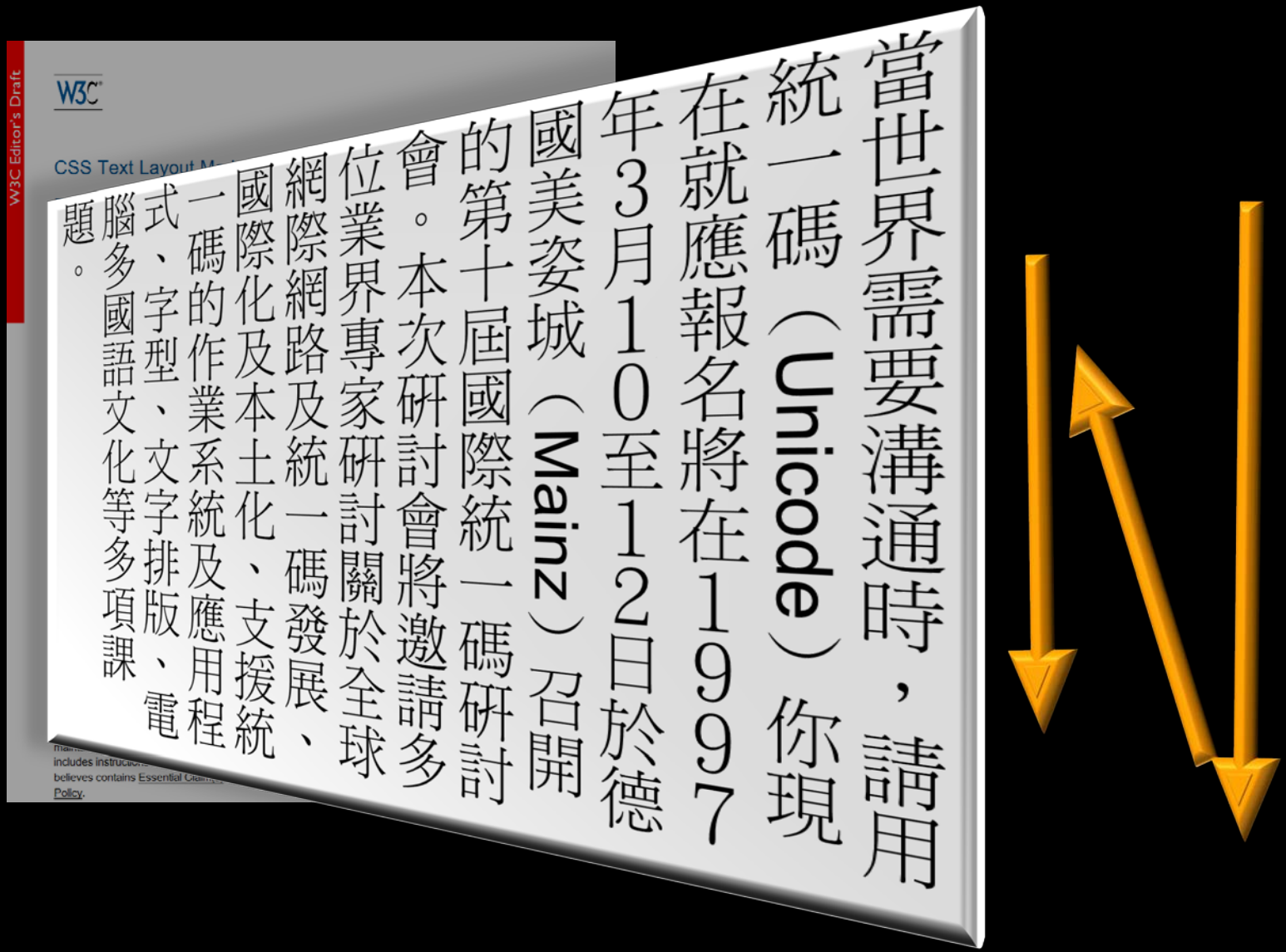
Измакоше ти кућу
па собу
па су ти узели свеску
из руке
неки бомбардери…

Беоīрад, Аīрил 1944

Измакоше īи кућу
īа собу
īа су īи узели свеску
из руке
неки бомбардери…

Standards support

Vertical text

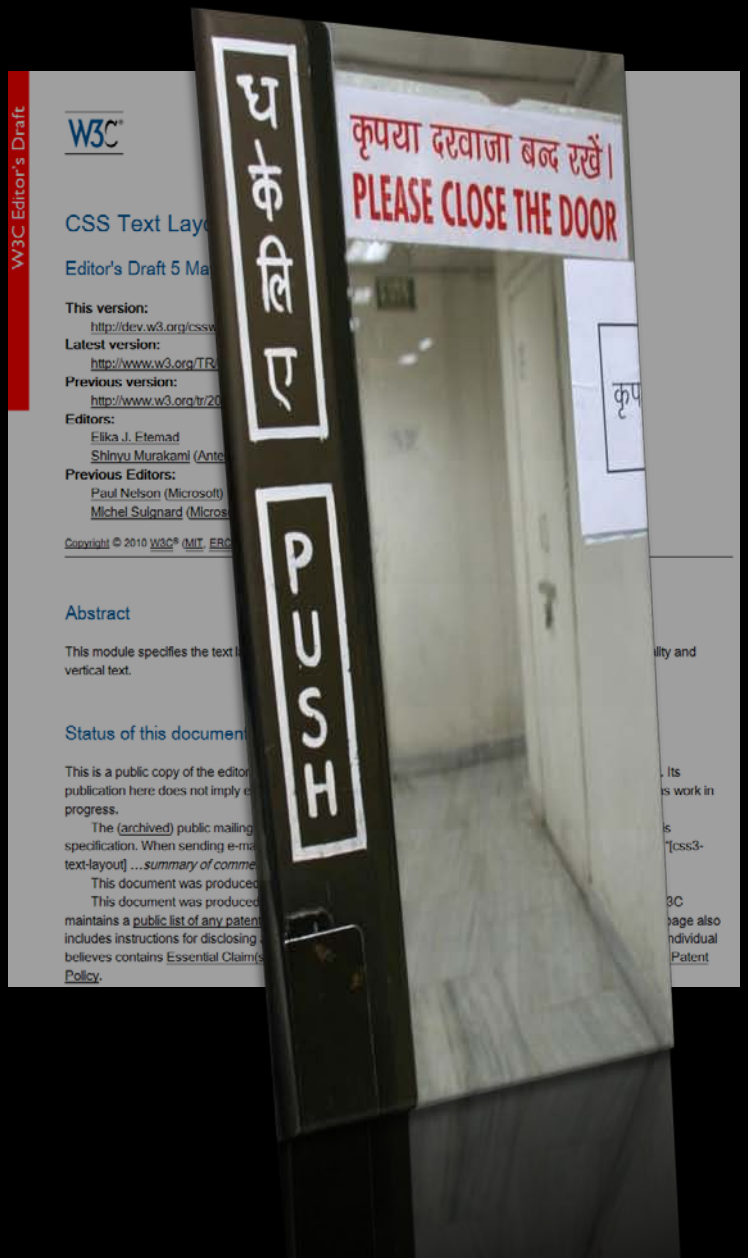


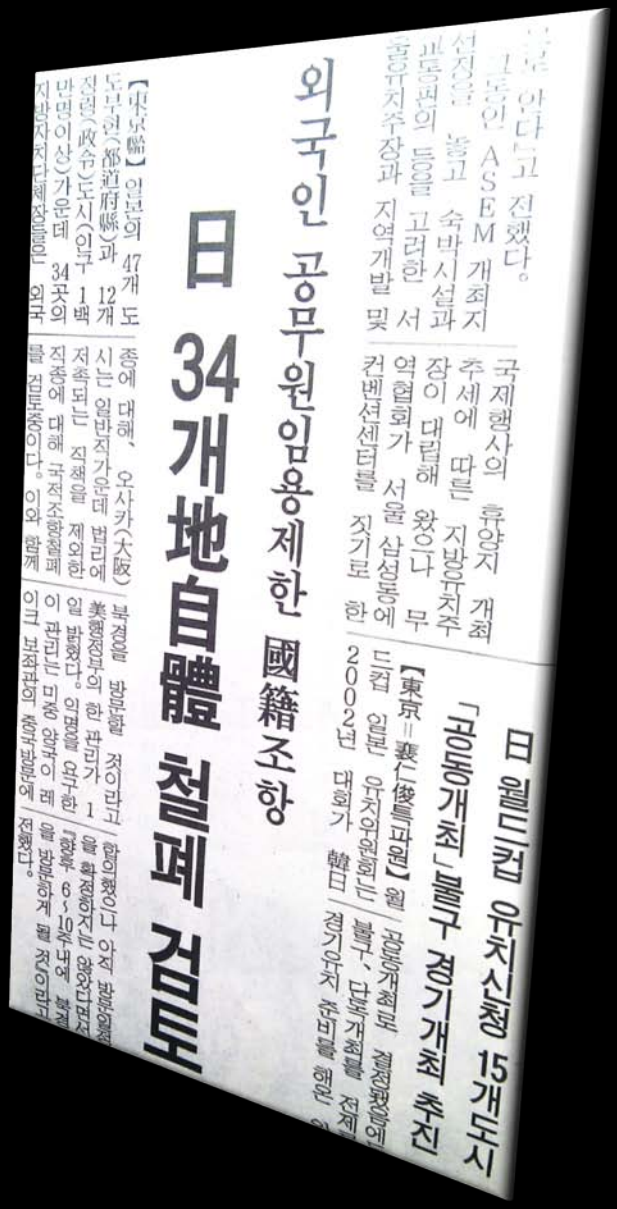
Standards support

Vertical text

Standards support

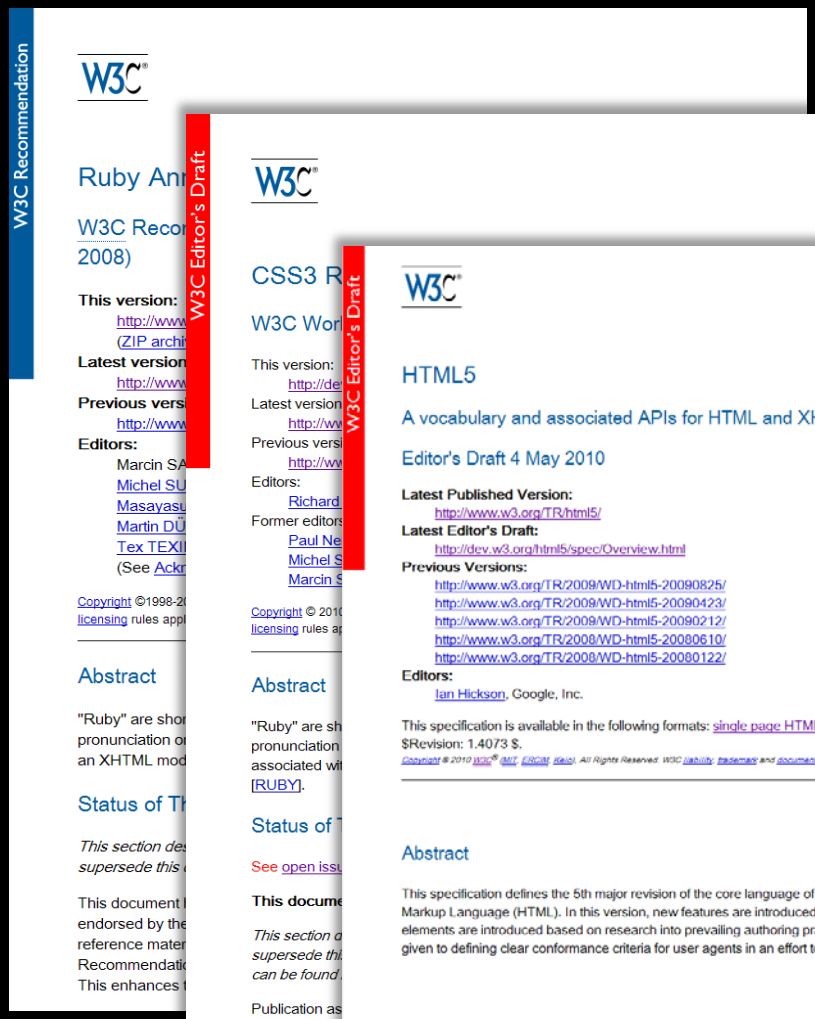
Vertical text

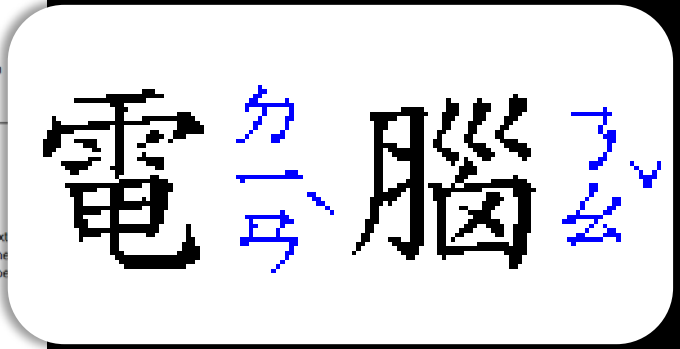




Standards support

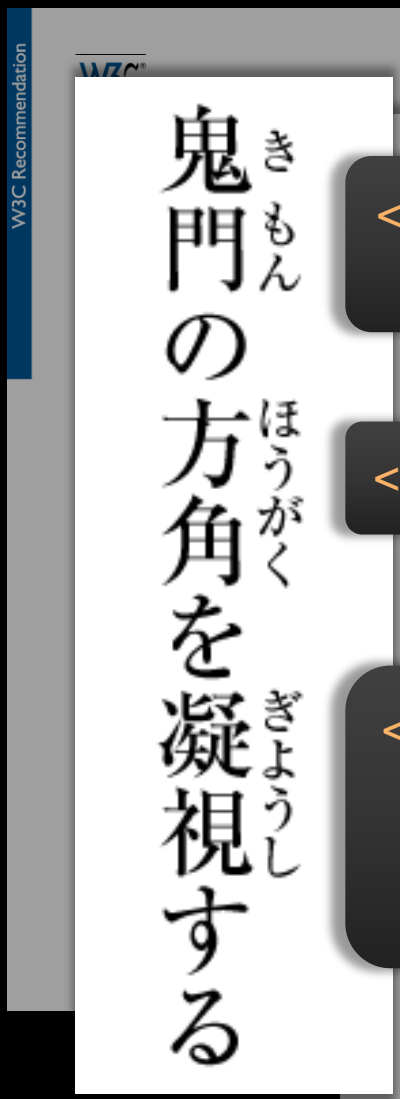
Ruby annotation





Standards support

Ruby annotation




```
<ruby<rb>凝</rb><rt>ぎょう</rt></ruby>
  <ruby><rb>視</rb><rt>し</rt></ruby>
```

HTML5

A vocabulary and associated APIs for HTML and XHTML

http://dev.w3.org/html5/spec/Overview.html

Previous Versions:

http://www.w3.org/TR/2009/WD-html5-20090825/
http://www.w3.org/TR/2009/WD-html5-20090423/
http://www.w3.org/TR/2009/WD-html5-20090212/
http://www.w3.org/TR/2008/WD-html5-20080610/
http://www.w3.org/TR/2008/WD-html5-20080122/

```
<ruby>凝<rt>ぎょう</rt>視<rt>し</rt></ruby>
```

```
<ruby>
  <rbc><rb>凝</rb><rb>視</rb></rbc>
  <rtc><rt>ぎょう</rt><rt>し</rt></rtc>
  </ruby>
```

Status of This document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of current W3C publications and the most recently formally published revision of this technical report can be found in the W3C technical reports index at http://www.w3.org/TR/.

If you wish to make comments regarding this document, please send them to public-html-comments@w3.org

Developing requirements

Requirements for Japanese Layout

W3C Working Group Note

[contents]

W3C

Requirements for Japanese Text La

W3C Working Group Note 4 June 2009

This version:
http://www.w3.org/TR/2009/NOTE-jlreq-20090604/
Latest version:
http://www.w3.org/TR/jlreq/
Previous version:
http://www.w3.org/TR/2008/WD-jlreq-20081015/
Editors:
Yasuhiro Anan (阿南 康宏), Microsoft
Hiroyuki Chiba (千葉 弘幸), Invited Expert
Junsaburo Edamoto (枝本 順三郎), Invited Expert
Richard Ishida, W3C
Keiichiro Ishino (石野 恵一郎), Antenna House
Tatsuo Kobayashi (小林 龍生), JustSystems
Toshi Kobayashi (小林 敏), Invited Expert
Kenzou Onozawa (小野澤 賢三), Invited Expert
Felix Sasaki, University of Applied Sciences Potsdam

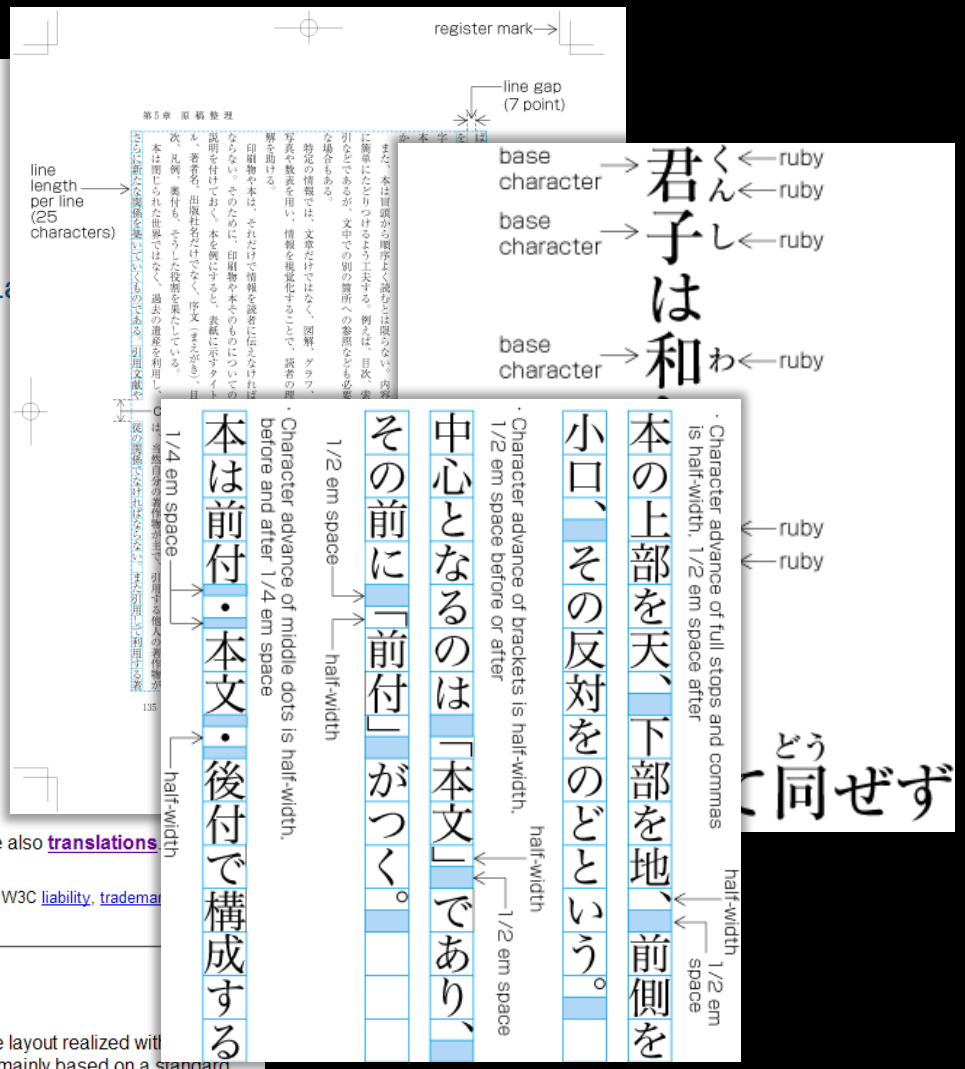
Please refer to the **errata** for this document.

A **Japanese version** of this document is also available. See also **translations**. English version of this document is the authoritative version.

Copyright © 2008-2009 W3C® (MIT, ERCIM, Keio), All Rights Reserved. W3C liability, trademar 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 standard for Japanese layout, JIS X 4051, however, it also addresses areas which are not covered by JIS X 4051.



Standards support

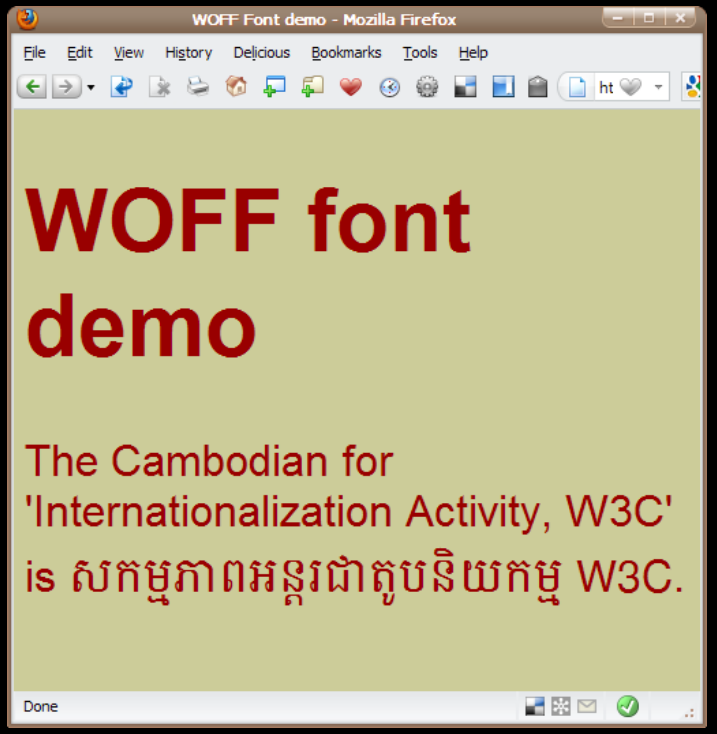
Web fonts

Standards support

Web fonts

```
@font-face {
  font-family: 'battambang-woff';
  font-style: normal;
  font-weight: normal;
  src: url(fonts/khmerosbbang.woff);
  }


:lang(kh) {
  font-family: 'battambang-woff';
  font-size: 100%;
  }
```

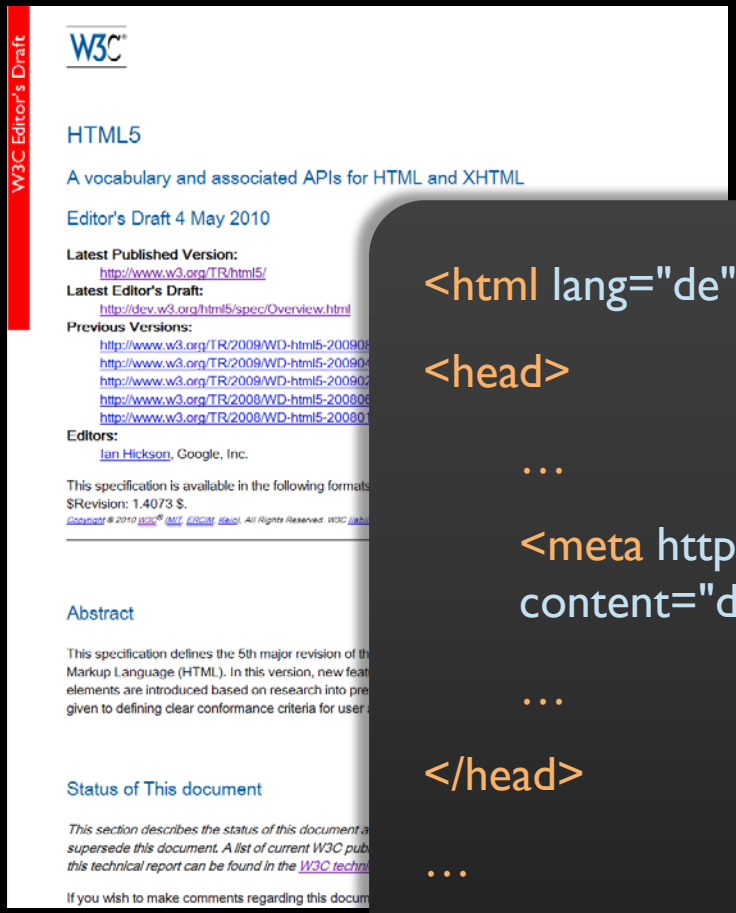


Issues

- Rendering detail for complex fonts.
- Subsetting capability may be needed.
- Can only be used for fonts with an appropriate licence.

Standards support

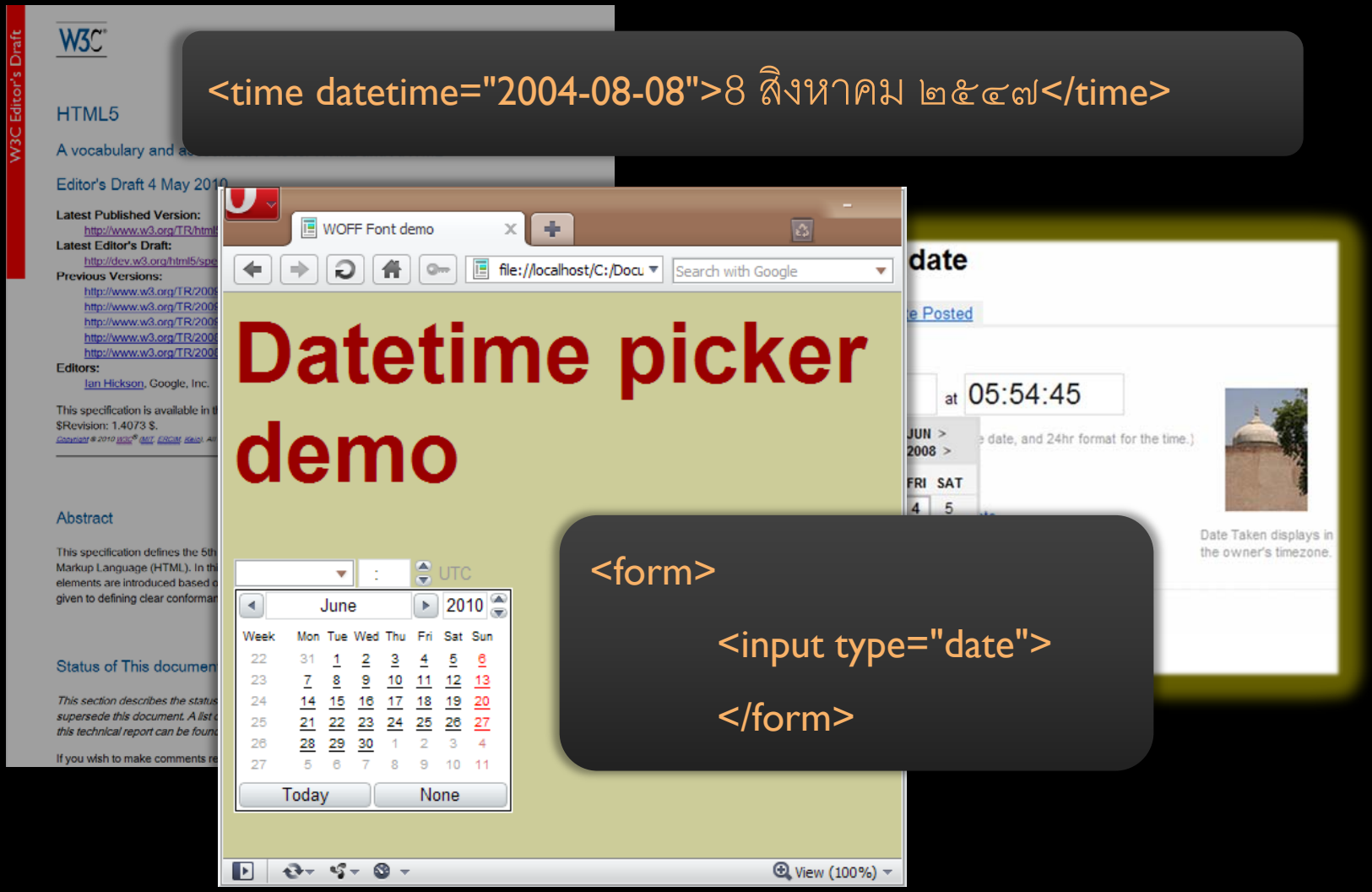
Language declarations in HTML5




```
<html lang="de">
<head>
    …
    <meta http-equiv="Content-Language"
    content="de">
    …
</head>
…
```

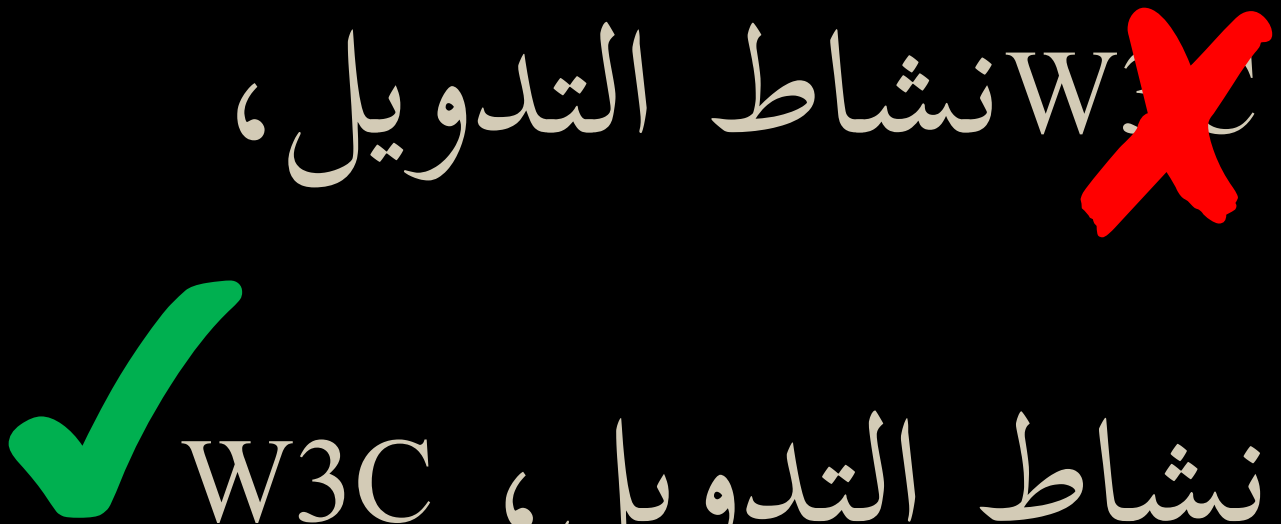
Standards support

Date and time



Standards support

Bidirectional text support

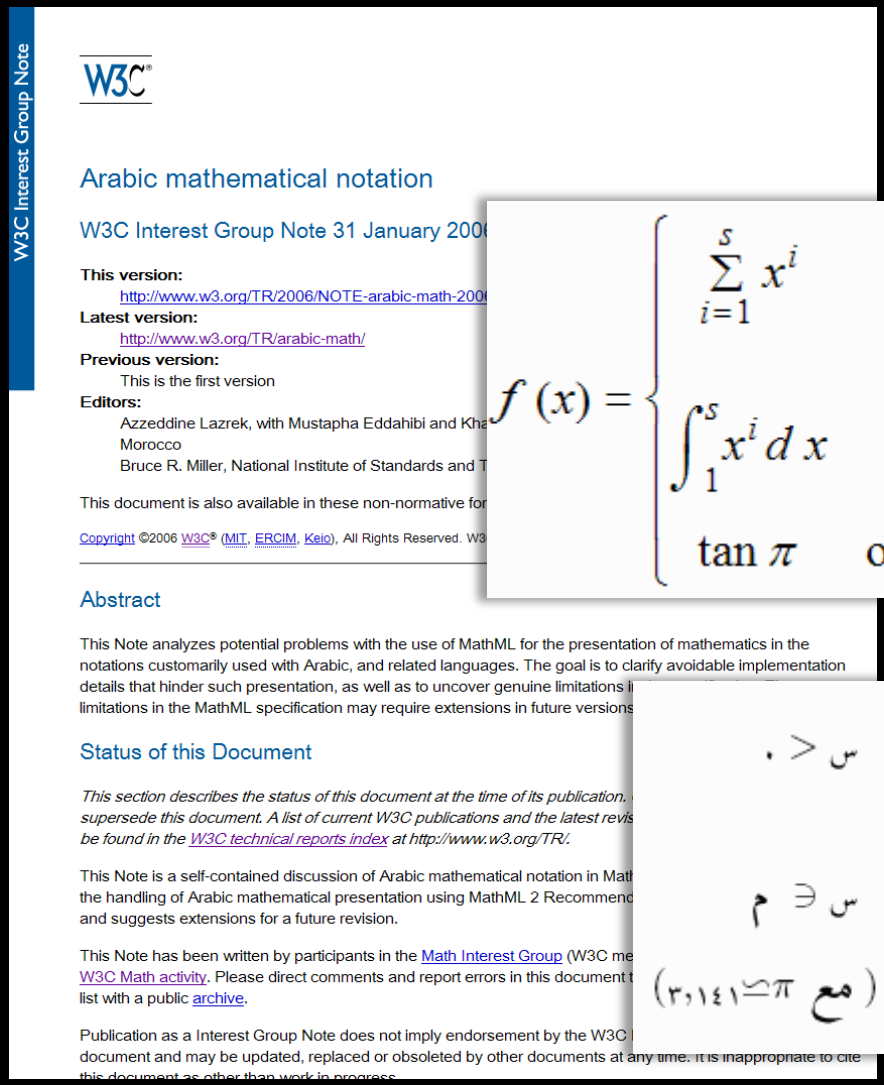



```
<description dir="rtl">W3C نشاط التدويل،</description>
```

Developing requirements

Augmenting bidi support in HTML5 & CSS

Developing requirements
Arabic mathematics



W3C Interest Group Note

W3C

Arabic mathematical notation

W3C Interest Group Note 31 January 200

This version:
http://www.w3.org/TR/2006/NOTE-arabic-math-200
Latest version:
http://www.w3.org/TR/arabic-math/
Previous version:
This is the first version
Editors:
Azzeddine Lazrek, with Mustapha Eddahibi and Kha
Morocco
Bruce R. Miller, National Institute of Standards and T

This document is also available in these non-normative for

Copyright ©2006 W3C® (MIT, ERCIM, Keio), All Rights Reserved. W3

$$f(x) = \begin{cases} \sum_{i=1}^{s} x^i & \text{if } x < 0 \\ \int_1^s x^i \, dx & \text{if } x \in S \\ \tan \pi & \text{otherwise (with } \pi \simeq 3.141) \end{cases}$$

Abstract

This Note analyzes potential problems with the use of MathML for the presentation of mathematics in the notations customarily used with Arabic, and related languages. The goal is to clarify avoidable implementation details that hinder such presentation, as well as to uncover genuine limitations i
limitations in the MathML specification may require extensions in future versions

Status of this Document

This section describes the status of this document at the time of its publication.
supersede this document. A list of current W3C publications and the latest revis
be found in the W3C technical reports index at http://www.w3.org/TR/.

This Note is a self-contained discussion of Arabic mathematical notation in Math
the handling of Arabic mathematical presentation using MathML 2 Recommend
and suggests extensions for a future revision.

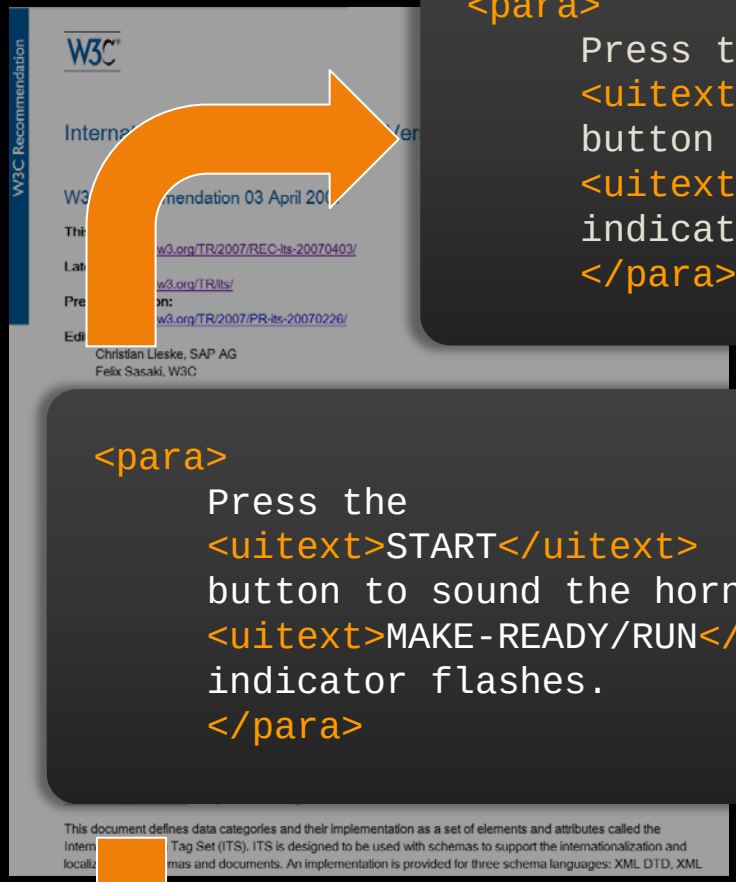
This Note has been written by participants in the Math Interest Group (W3C me
W3C Math activity. Please direct comments and report errors in this document t
list with a public archive.

Publication as a Interest Group Note does not imply endorsement by the W3C
document and may be updated, replaced or obsoleted by other documents at any time. It is inappropriate to cite
this document as other than work in progress.

$$\text{ت}(\text{س}) = \begin{cases} \text{مجـ}_{\text{ب}=١}^{\text{ص}} \text{س}^{\text{ب}} & \text{إذا كان } \text{س} < ٠ \\ \int_{١}^{\text{ص}} \text{س}^{\text{ب}} \text{، س} & \text{إذا كان } \text{س} \in \text{م} \\ \text{ظا } \pi & \text{غير ذلك (مع } \pi \simeq ٣{,}١٤١) \end{cases}$$

Standards support

Internationalization Tag Set




```
<para>
      Press the
      <uitext translate="no">START</uitext>
      button to sound the horn. The
      <uitext translate="no">MAKE-READY/RUN</uitext>
      indicator flashes.
      </para>
```

```
<para>
      Press the
      <uitext>START</uitext>
      button to sound the horn. The
      <uitext>MAKE-READY/RUN</uitext>
      indicator flashes.
      </para>
```

- supported by some translation tools – linked with XLIFF
- being applied by specifications at W3C

```
<its:rules ... its:version="1.0">
    <its:translateRule selector="//uitext" translate="no"/>
    </its:rules>
```

The changing social context

Social context

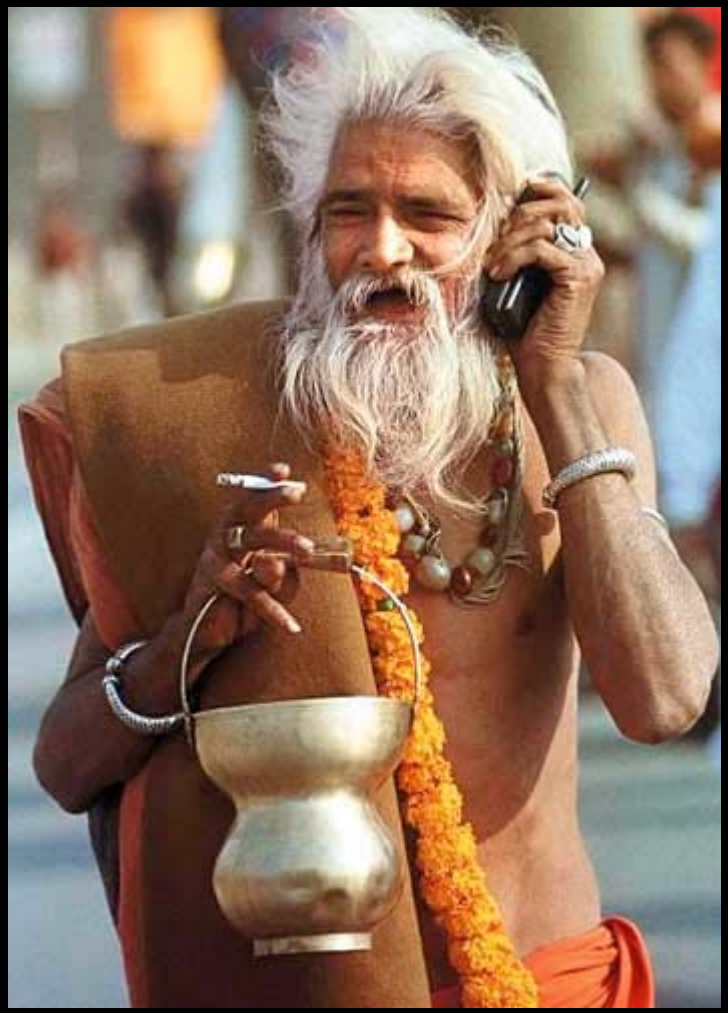
The rise of the Mobile Web

- "In China … over 73m people, or 29% of all internet users in the country, use mobile phones to get online."
- "The number of pages viewed in June by 14m users of [Opera] software was over 3 billion, a 300% increase on a year earlier. The fastest growth was in developing countries including Russia, Indonesia, India and South Africa."

Economist.com, Sept. 2008

Social context

Mobile Web for Developing Society (MW4D)

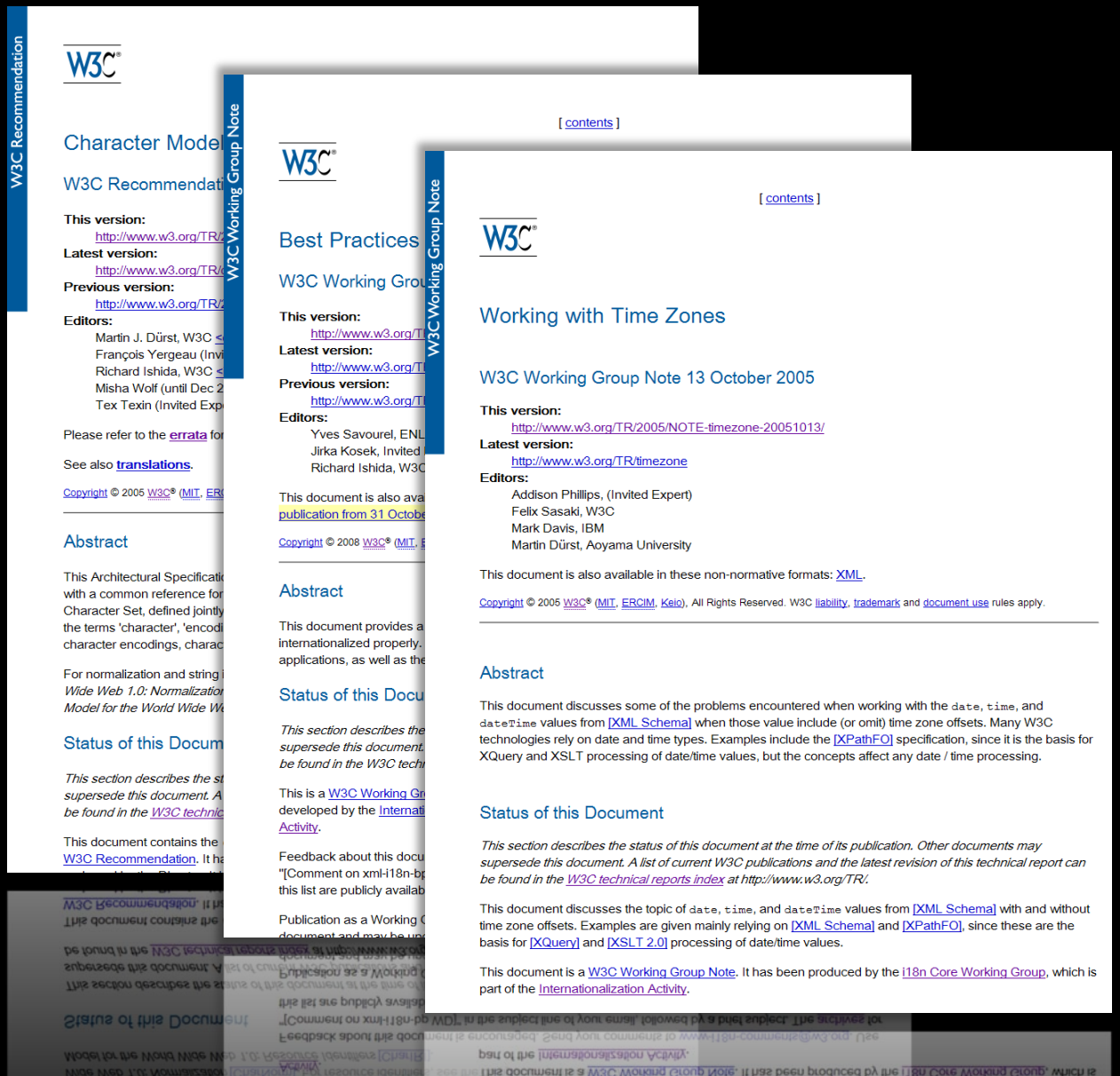


Track the social impact of the mobile web in the developing world, to ensure that the web's technical standards evolve to serve this rapidly emerging constituency.

Best practices for the multilingual Web

Best practices

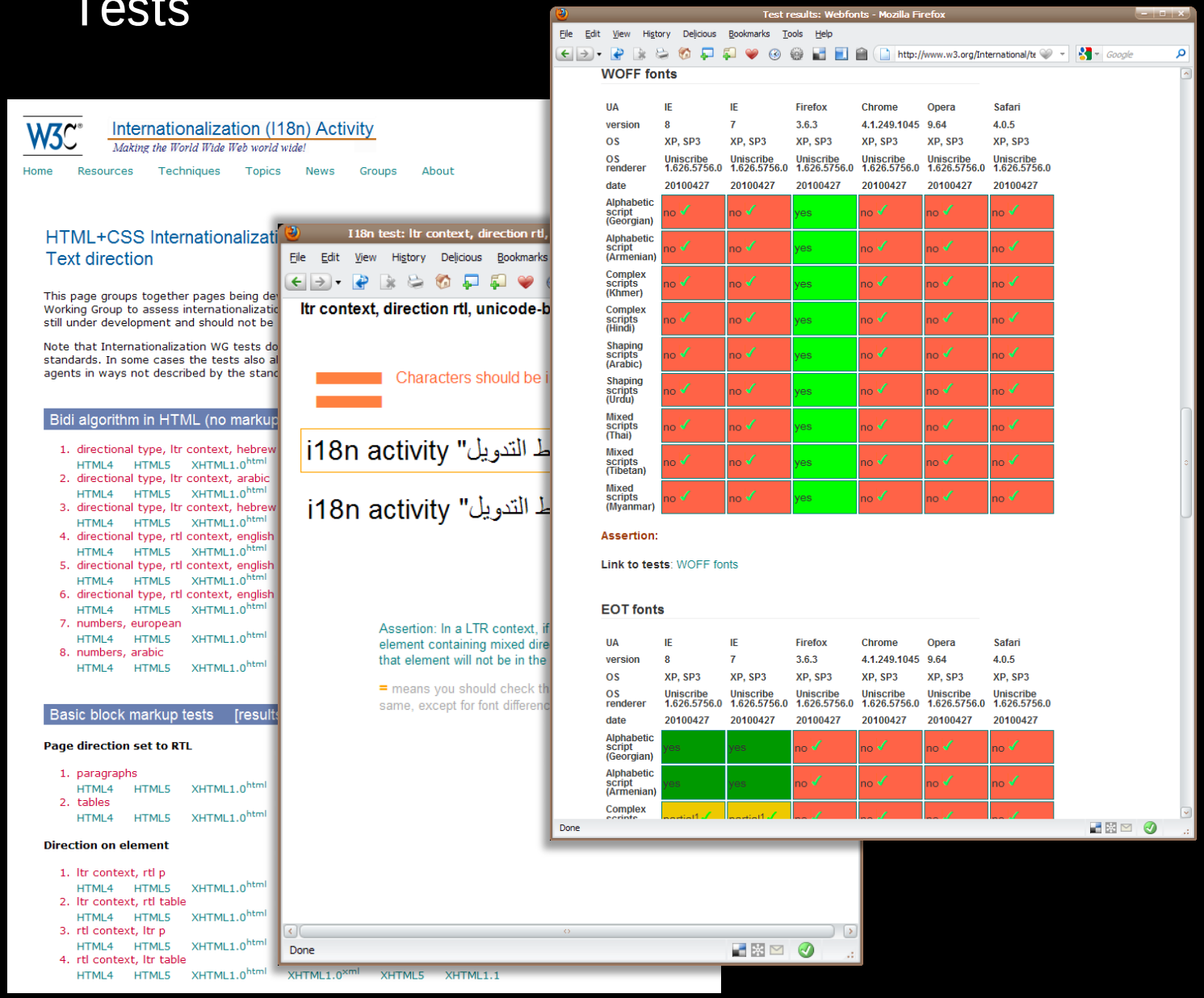
Capturing guidance for spec developers



- markup for bidirectional text
- Normalization
- working with case sensitivity
- more information about date & time

Best practices

Tests

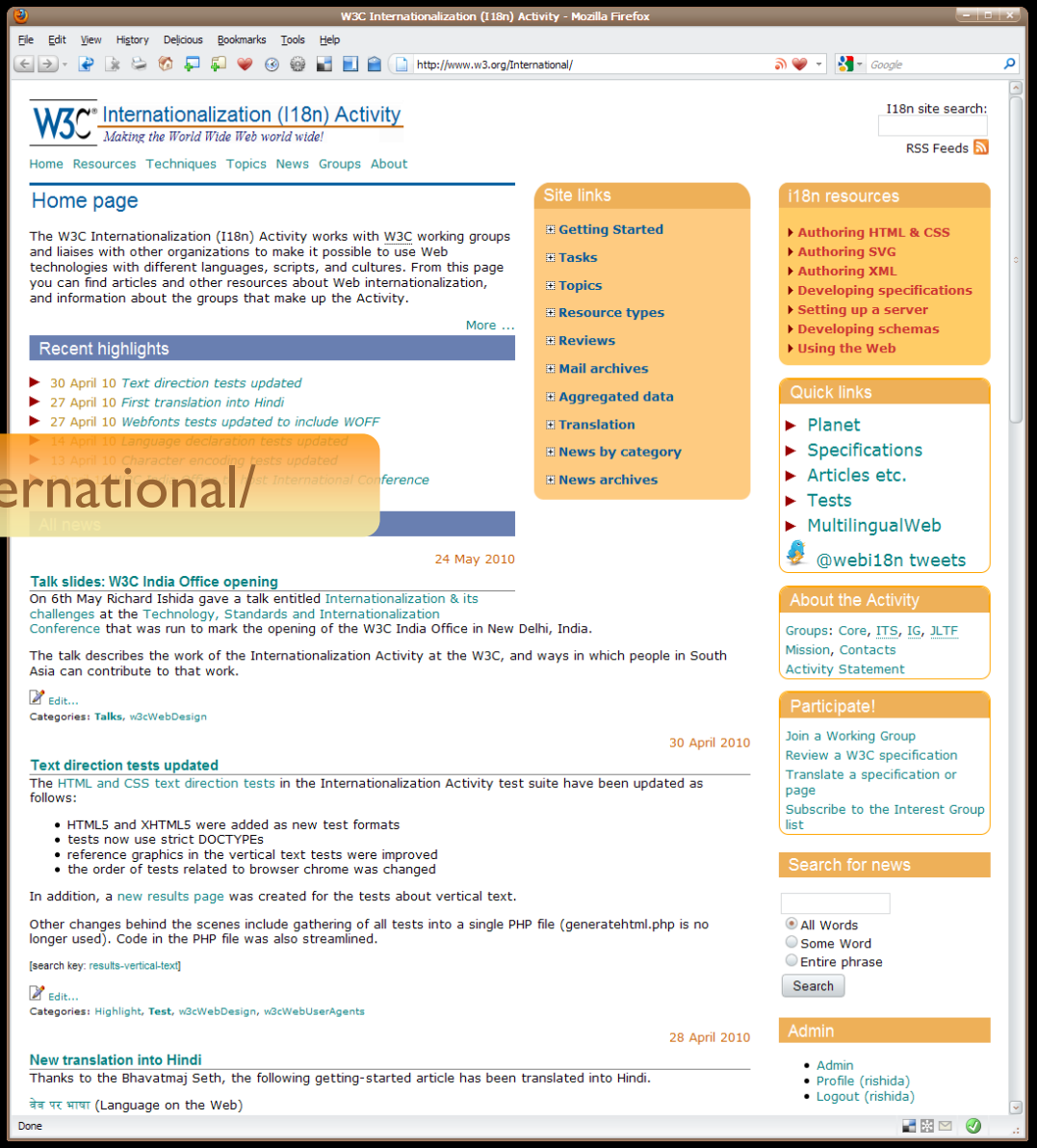


Best practices

I18n resources

- Articles
- Tutorials
- Technical notes
- Tests
- Talks
- Tools
- Reviews

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



Best practices

I18n resources

- Articles
- Tutorials
- Technical notes
- Tests
- Talks
- Tools
- Reviews

W3C Internationalization (I18n) Activity

Making the World Wide Web world wide!

I18n site search:

RSS Feeds

Home Resources Techniques Topics News Groups About

Articles, best practices & tutorials

You can also find resources using the Technique index and Topic index, which provide more fine-grained access to information.

Getting Started

- Overview
- Introducing character sets and encodings
- Language on the Web
- Internationalization Quick Tips for the Web

Characters

- Character encodings for beginners
- Character encodings
- Character sets & encodings in XHTML, HTML and CSS
- Changing (X)HTML page encoding to UTF-8
- Setting encoding in web authoring applications
- Using character entities and NCRs
- Document character set
- CSS character encoding declarations
- Setting the HTTP Charset parameter
- Setting charset information in .htaccess
- Checking HTTP headers
- Checking the character encoding using the validator
- Character Model for the World Wide Web 1.0: Fundamentals
- Display problems caused by the UTF-8 BOM
- HTML, XHTML, XML and control codes
- Missing characters and glyphs
- Who uses Unicode?
- Migrating to Unicode

Language

- Specifying Language in XHTML & HTML Content
- Language tags in HTML and XML
- Choosing a language tag NEW!
- 2-letter or 3-letter language codes
- Why use the language attribute?
- Setting language preferences in a browser
- Declaring Language in XHTML and HTML
- xml:lang in XML document schemas

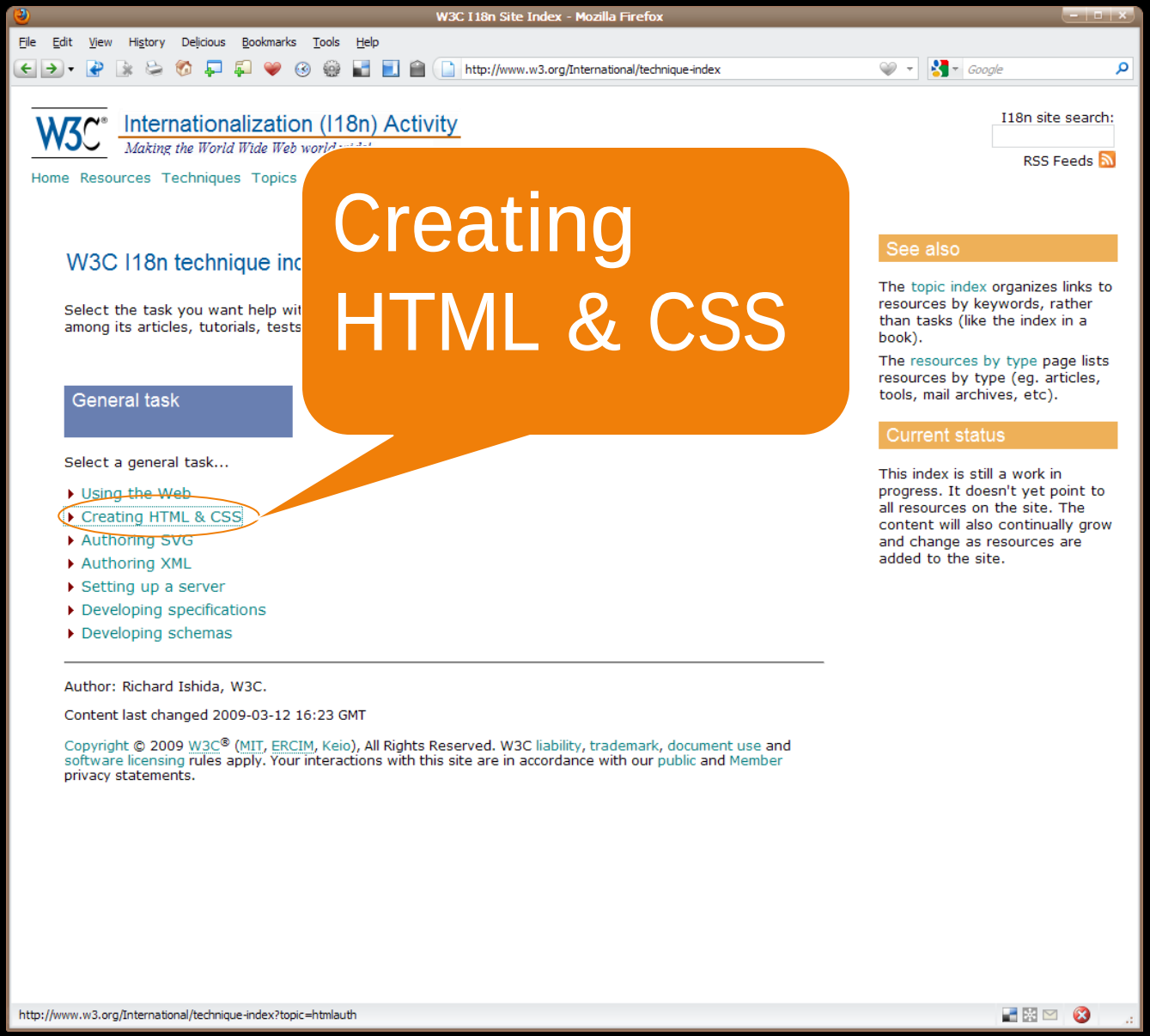
> عربي Български Deutsch Ελληνικά Español Français עברית हिन्दी Magyar Italiano 日本語 한국어 Nederlands Polski Português Português-BR Română Русский Svenska ไทย Türkçe Українська Tiếng Anh 简体汉语 繁體中文

On this page

- Getting started
- Characters
- Language
- Markup & text
- Text direction
- Styling & layout
- Forms
- Navigation
- Web addresses
- Cultural issues
- Other

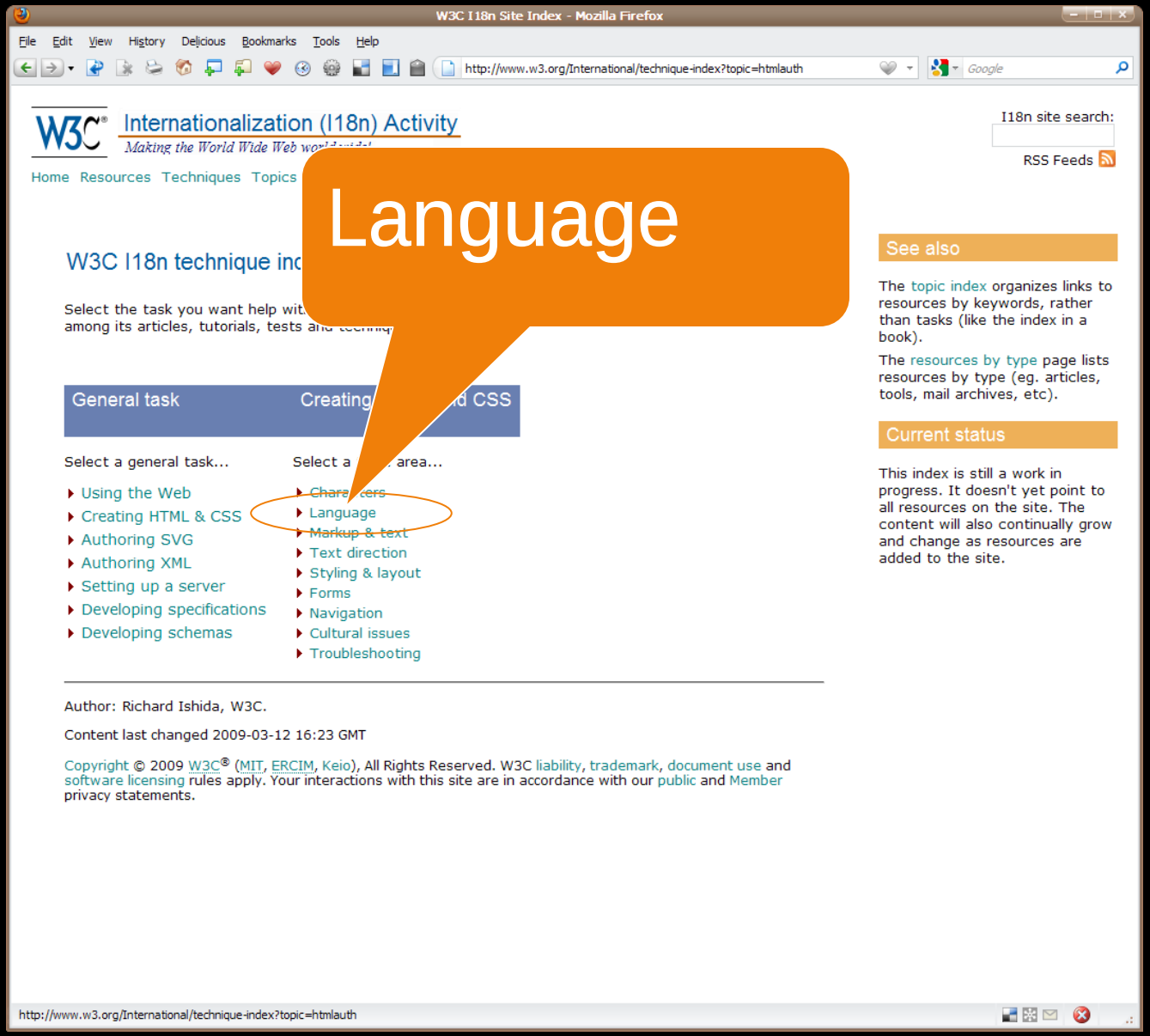
Best practices

I18n resources



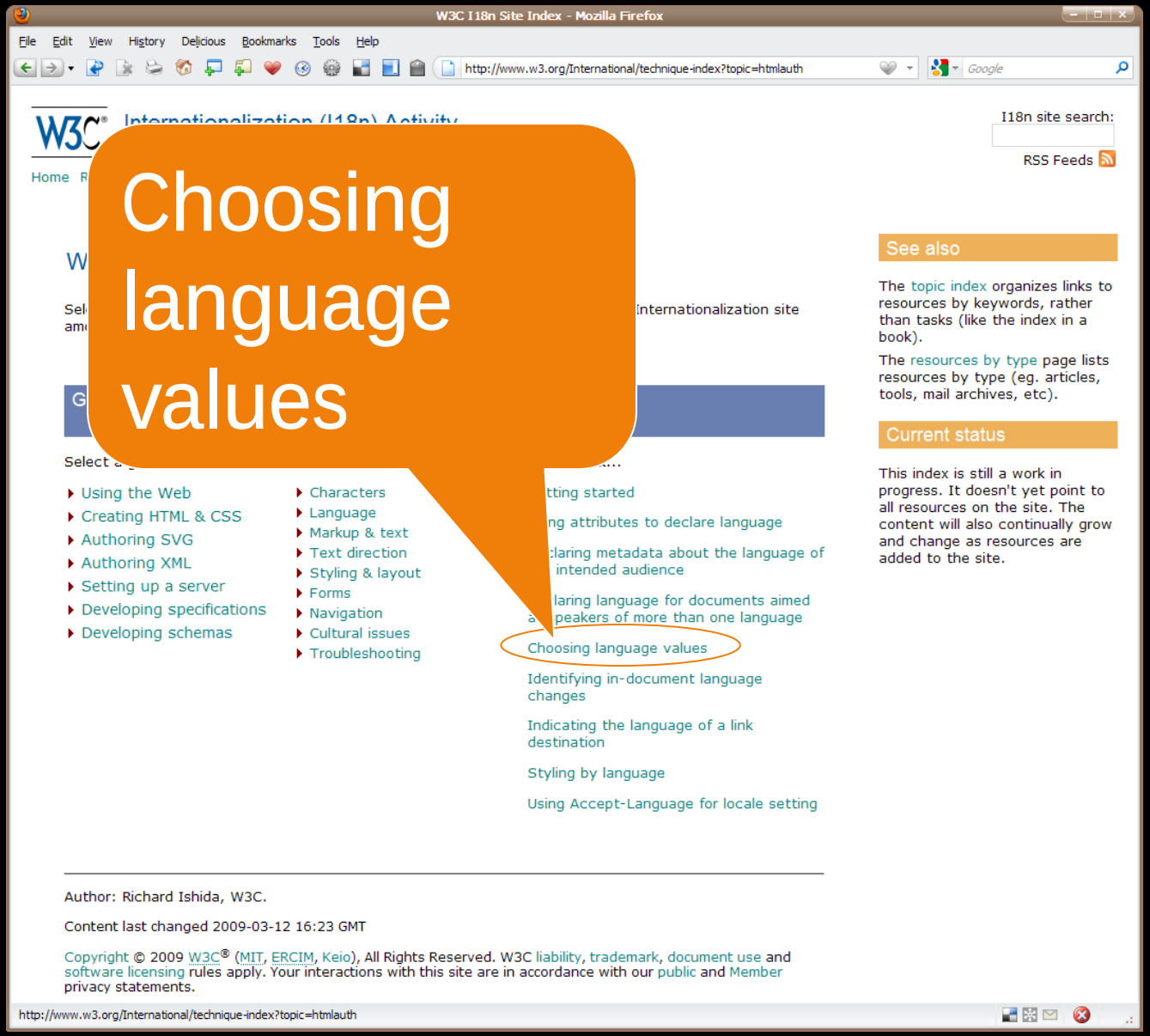
Best practices

I18n resources



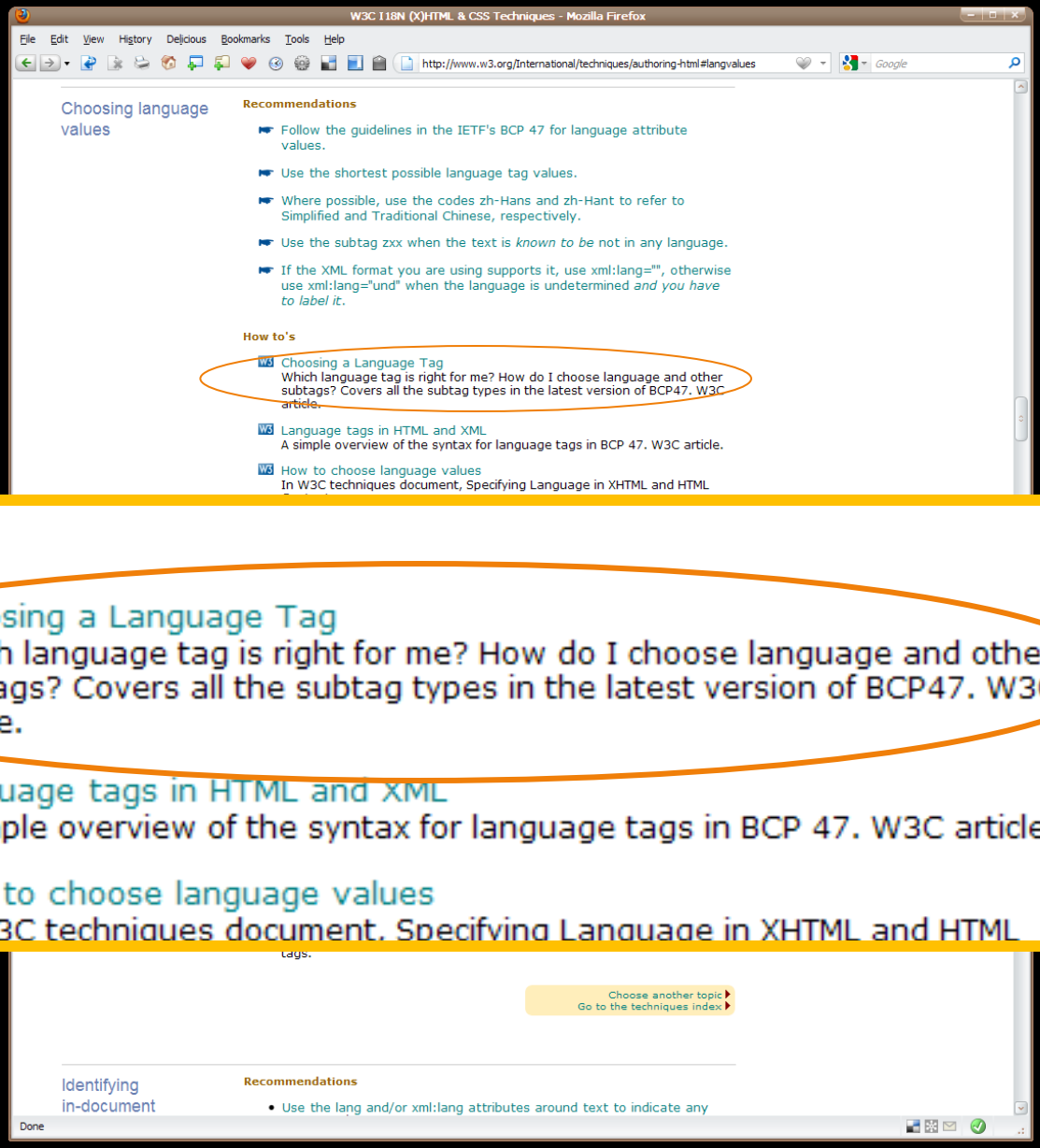
Best practices

I18n resources



Best practices

I18n resources



How to's

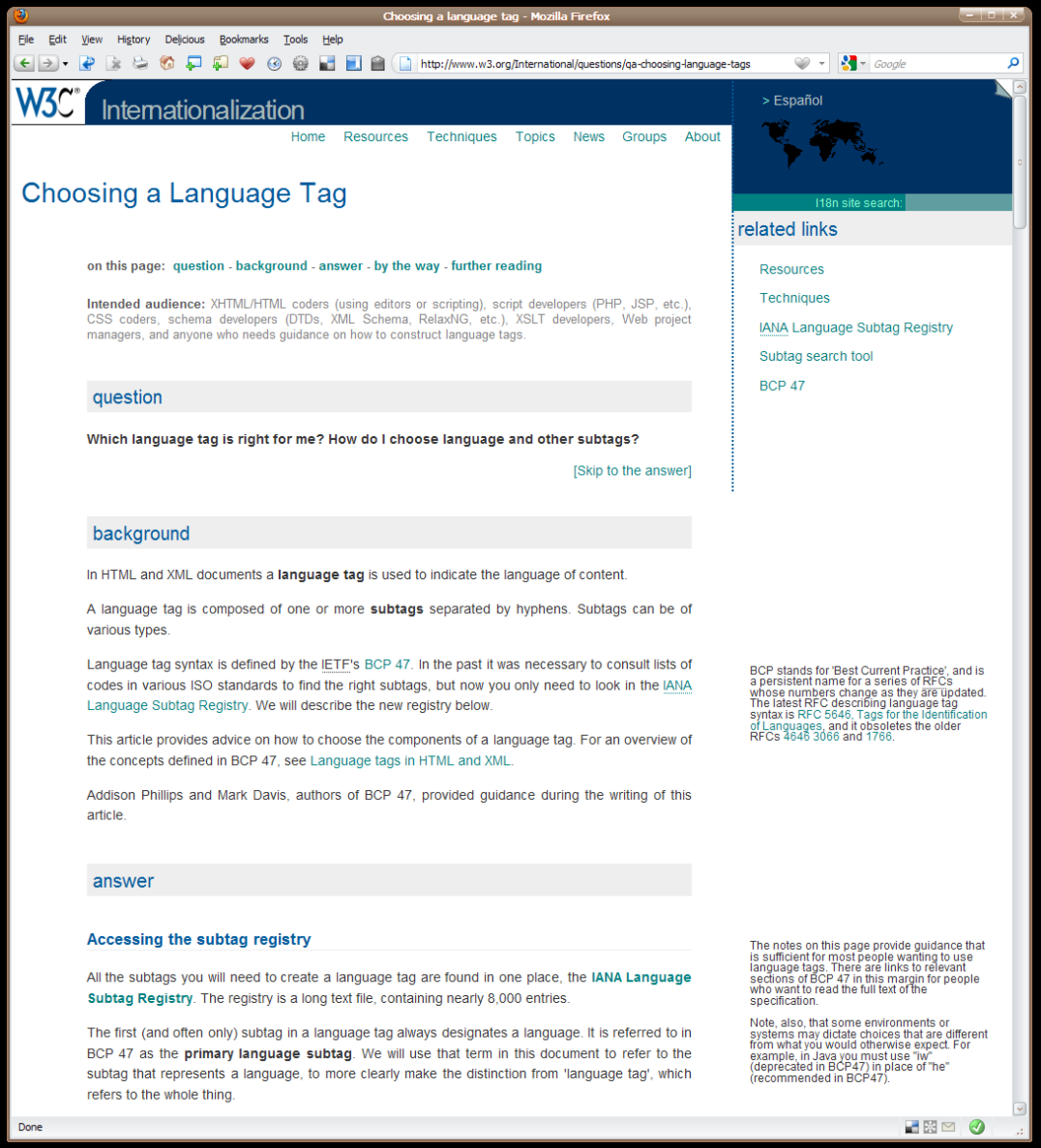
Choosing a Language Tag
Which language tag is right for me? How do I choose language and other subtags? Covers all the subtag types in the latest version of BCP47. W3C article.

Language tags in HTML and XML
A simple overview of the syntax for language tags in BCP 47. W3C article.

How to choose language values
In W3C techniques document, Specifying Language in XHTML and HTML

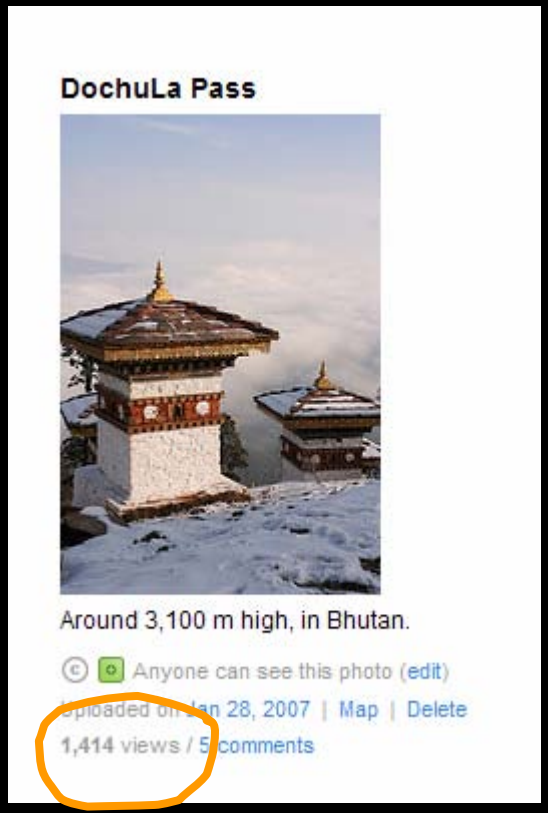
Best practices

I18n resources



Best practices

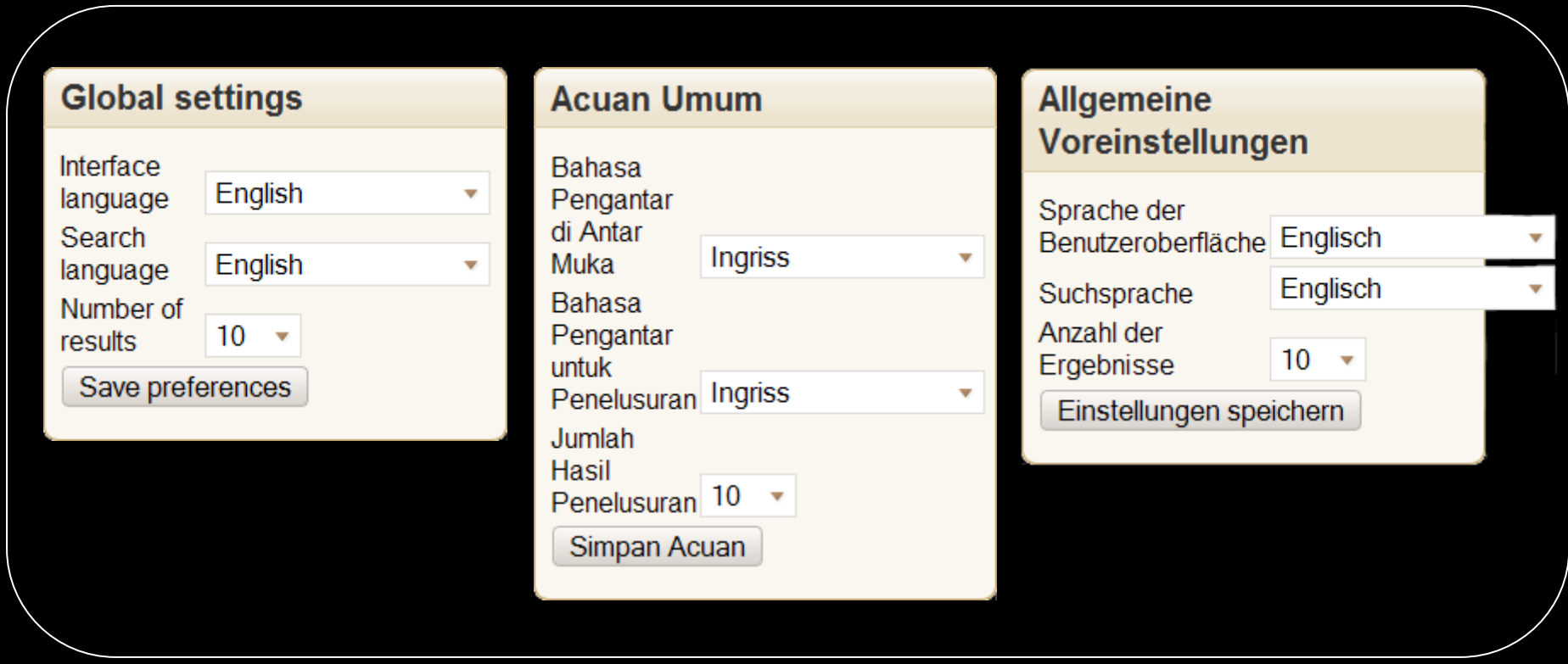
Text expansion



Korean		0.8
English	views	1
Chinese	次檢視	1.2
Portuguese	visualizações	2.6
French	consultations	2.6
German	-mal angesehen	2.8
Italian	visualizzazioni	3

Best practices

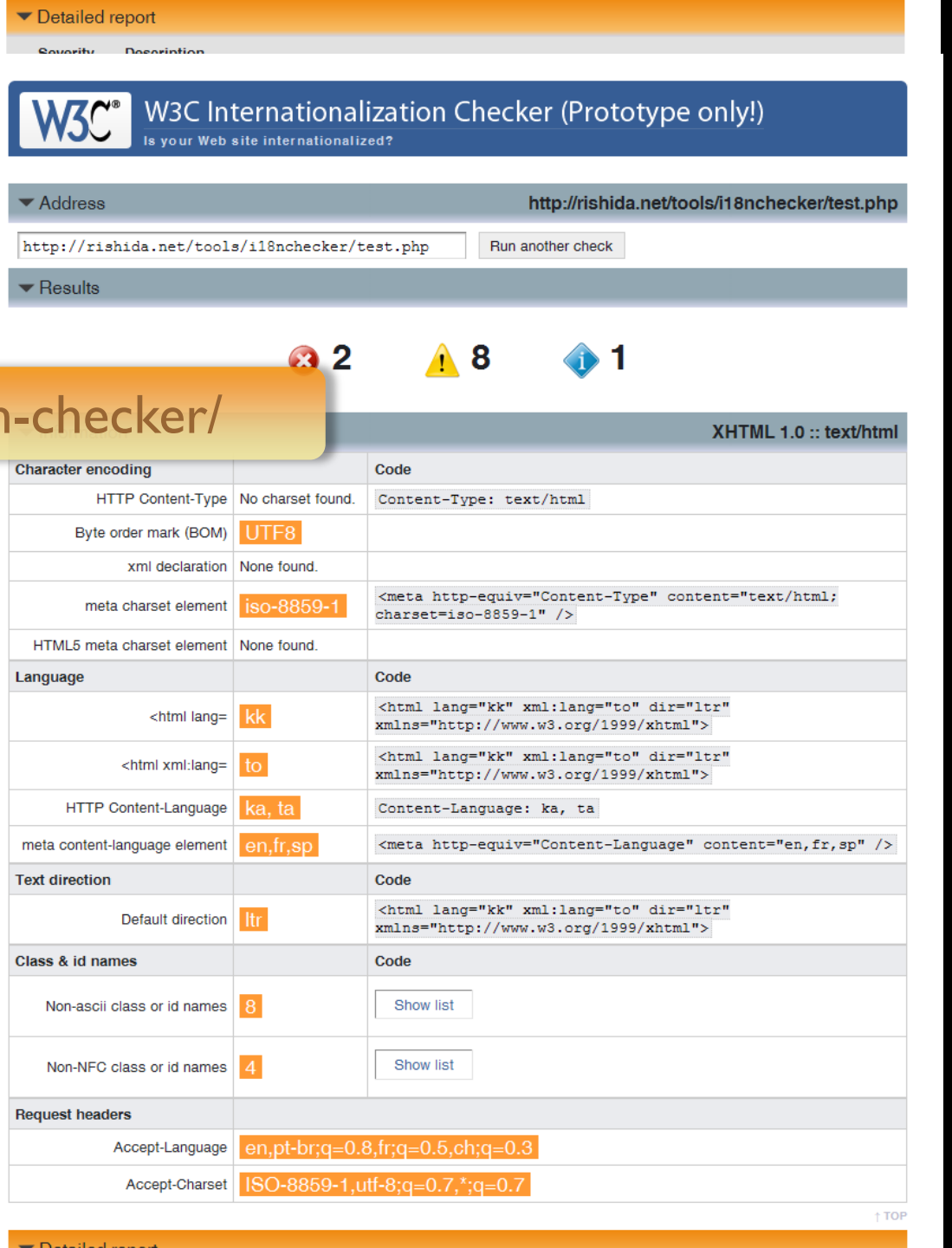
Text expansion



Best practices
Checker tool

http://validator.w3.org/i18n-checker/

1. Discover
2. Check



Getting involved…

Getting involved

- Follow the discussions on the i18n mailing lists (eg. www-international@w3.org), and track other technologies for internationally relevant topics. Follow our RSS feeds and twitter channels (@webi18n and @multilingweb)
- Read and review specifications (http://www.w3.org/TR/tr-technology-drafts) and send comments to the i18n list or direct to the Working Group.
- Discuss local requirements for the Multilingual Web, and if you identify missing features, find ways to coordinate proposals.
- Use features needed for non-Latin script support and push implementers to include more in browsers and authoring tools.

Getting involved

- Review or contribute to development/dissemination of outreach materials, to help others understand how to implement and use international features of the Web.
- Take on board that internationalization is something done by developers and designers – not localizers. Find out how to do it. (http://www.org/International/)
- Use Unicode wherever you can.
- Consider how your content will appear on the Mobile Web.
- Participate in the MultilingualWeb events planned over the coming year and a half.
- Use the I18n Checker (http://qa-dev.w3.org/i18n-checker/) and send ideas for improvements.
- Don't rely on us to do the work for you! We need your help.

The Web needs your help

this is your Web –
not the W3C's

the Web is about
people, not
technology

we need You to
make the Web
worldwide

get involved

Thank you

http://www.w3.org/International/talks/1010-madrid/