Outline

About the W3C
Standards support for the multilingual Web
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


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

Unicode

جعل شبكة الويب العالمية عالمية حقًا!
وبظلال رادار سي جالاك مازم!
عالم الويب كوكب طوره عالميًا بالمغرب

"دَنيَةِ ُعيِّزِلِيكَِ يُودِّ" هَاغَيَ دُنيَعِجُعْزِلِيكَِ يُتِمِّيَزْ!

缔造真正全球通行的万维网

Gwneud y we fyd-eang yn wirioneddoll fyd-eang!
Standards support Unicode

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

Tim Berners-Lee
Standards support

Unicode

جعل شبكة الويب العالمية حقًا! 
وبسائل رأسي فعلاً ما زلنا,
عالمگیر ویب کو حقیقی طور پر عالمگیر بنانای

Unicode on the Web
ჭიამაია (Coccinellidae), ხოჭოების ოჯახს ეკუთვნის. აქვს ამობურცული, მომრგვალო ან ოვალური სხეული. ზურგზე ღია ფონზე შავი ლაქები აყრია, იშვიათად...
ჭიამაია (Coccinellidae), ხოჭოების სურათის სიარულით ეკუთვნის. მალაგარი ამობურცული, მომრგვალო ან დარგული სხეული. ზურგზე ღია ფონზე შავი ლაქები აყრია, იშვიათად ...
Ha a világ beszélni akarna, Unicode-ul szólalna meg. Regisztráljon már most a Tizedik Nemzetközi Unicode Konferenciára, melyet 1997. március 10-12-én rendeznek Meinz-ban, Németországban. Ezen a konferencián az iparág több neves szakértője is résztvesz. Ízelítőül a témákból: a világháló és a Unicode nemzetközisítése és lokalizálása, a Unicode alkalmazása működő rendszerekben és alkalmazásokban, szövegelrendezésnél, és többyelvű számítógépeken.
Standards support

Web resource identifiers


Scheme
Domain name
Path

xn--jp-cd2fp15c.xn--fsq.jp
Standards support
Top level domain names

السعودية
أمانات
مصر

Al-Saudiah
Emarat
Misr

وزارة-الاتصالات.مصر/
Standards support
Web resource identifiers

Scheme: http
Domain name: 納豆.例.jp/dir1/
Path: 引き割り.html

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

Language tags

- ISO 639 language codes
- ISO 3166 country codes

Before (RFC 3066)

en
en-GB
en-scouse
Standards support

Language tags: BCP 47

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

Now BCP 47

language script region variant extension private_use
(extlang)

hi az-Cyrl zh-Hans es-419 sl-IT-rozaj-njiva-1994
Standards support

Key Events

Document Object Model (DOM) Level 3 Events Specification
W3C Editor's Draft 21 April 2010

This version:
http://dev.w3.org/2008/webapi/DOM-Level-3-Events.html

Latest stable version:
http://www.w3.org/TR/DOM-Level-3-Events

Previous version:

Editor's Draft:
http://dev.w3.org/2006/webapi/DOM-Level-3-Events/html/DOM3-Events.html

Editors:
Doug Schepers, W3C
Björn Höflehner, Invited Expert (until December 2007)
Philippe Le Hégarat, W3C (until November 2003)
Toni Pizza, Netscape Communications Corporation (until July 2002)

Copyright © 2009 W3C (Attribution-NonCommercial-NoDerivs License).

Abstract

This specification defines the Event Model platform- and language-neutral models describing event flow through event for each event. The Document Object Model (DOM) is a platform and language-neutral model for representing and accessing the content and structure of documents. The Document Object Model Event Model is the set of event objects and to which the events that can be fired. The Event Model consists of an API for creating event objects and the specification of a set of events that can be fired.

Internationalization Quicktips

- Use Unicode wherever possible for content, databases, etc. Always declare the encoding of content.
- Use characters rather than escapes (e.g. \u{1E1} &\u{225}; or &\uacute;) whenever you can.
- Declare the language of documents and indicate internal language changes.
Developing requirements
Speech Synthesis Markup Language

Speech Synthesis Markup Language
W3C Proposed Recommendation 23 February 2010

This version:
http://www.w3.org/TR/2010/PR-speech-synthesis11-20100223/

Latest version:
http://www.w3.org/TR/speech-synthesis11/

Previous version:

Editors:
Daniel C. Burnett, Voxeo (formerly of Vocalocity and Nuance)
双喜伟 (Zhi Wei Shuang), IBM

Authors:
Paolo Baglio, Log语音
Paul Bagshaw, France Telecom
Michael Bodell, Microsoft
董方 (Hu Fang), Invited Expert (until 20 October 2009)
戴德普 (Zhi Huang), France Telecom
黄力行 (Li Xing Huang), Chinese Academy of Sciences
陈永国 (Yongguo Kang), Panasonic Corporation (until 5 December 2007)
林小辉 (Lin Xiaoqiang), Toshiba
Scott McGlashan, HP
蔡其荣 (Helen Meng), Chinese University of Hong Kong (until 29 July 2009)
陶镇华 (Jianhua Tao), Chinese Academy of Sciences
王霞 (Wang Xia), Nokia (until 90 October 2008)
夏海明 (Xia Haiming), Panasonic Corporation (until 2 August 2009)
吴志勇 (Zhiyong Wu), Chinese University of Hong Kong (until 29 July 2009)
严祝 (Yan Jun), FLYTEK

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.
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.

* { 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…

Београд, Април 1944
Измакоше Џи кућу Џа собу
Џа су Џи узели свеску из руке неки бомбардери…
當世界需要溝通時，請用統一碼（Unicode）。

你現在就應報名將在1997年3月10至12日於德國美姿城（Mainz）召開的第十屆國際統一碼研討會。本次研討會將邀請多位業界專家研討關於全球化、國際化及本土化、支援統一碼的作業系統及應用程式的字型、文字排版、電腦多國語文化等多項課題。
Standards support
Standards support

Vertical text
Standards support
Ruby annotation

HTML5
A vocabulary and associated APIs for HTML and XHTML

Editor's Draft 4 May 2010
Latest Published Version:
http://www.w3.org/TR/html5/
Latest Editor's Draft:
http://dev.w3.org/html5/Elements/
Previous Versions:
http://www.w3.org/TR/2006/WD/html5-20060111/
http://www.w3.org/TR/2005/WD/html5-20051215/
http://www.w3.org/TR/2004/CR-xml-20040913/
http://www.w3.org/TR/2004/CR-xml-20040621/

Editors:
Anne Hering, Google, Inc.

This specification is available in the following formats:
plain-text: HTML, xml/sgml-HTML. This is revision
SProcem 1.4.07/5 S

Abstract
This specification defines the 5th major revision of the core language of
the World Wide Web: the HyperText Markup Language (HTML). In this
version, new features are introduced to help Web application authors,
such as microformats, semantic HTML, and new APIs for more
utility. The changes were based on feedback from the W3C
working group, and the public.

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 at the W3C technical report index at
http://www.w3.org/TR/
Standards support

Ruby annotation

HTML5

A vocabulary and associated APIs for HTML and XHTML

Previous Versions:
http://www.w3.org/TR/2008/WD/html5-20081028/
http://www.w3.org/TR/2006/WD/html5-20080615/
http://www.w3.org/TR/2006/WD/html5-20080612/
http://www.w3.org/TR/2006/WD/html5-20080610/
http://www.w3.org/2006/06/html5/

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

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 https://www.w3.org/TR/
Developing requirements
Requirements for Japanese Layout

W3C Working Group Note 4 June 2009

This version:
http://www.w3.org/TR/2009/NOTE-jireq-20090604/
Latest version:
http://www.w3.org/TR/jireq/
Previous version:
http://www.w3.org/TR/2008/WD-jireq-20081015/

Editors:
Yasuhiro Anan (岡村 一宏), Microsoft
Hiroyuki Chiba (千葉 弘幸), Invited Expert
Jun-ichiro Edamoto (佐藤 一彦), Invited Expert
Richard Ishida, W3C
Keigo Ishino (石野 淑一郎), Antenna House
Tatsuo Kobayashi (小林 素之), JustSystems
Toshi Kobayashi (小林敏), Invited Expert
Kenzo Ono (小野 浩), 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.

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

```css
@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.
Language declarations in HTML5

Standards support

HTML5

A vocabulary and associated APIs for HTML and XHTML

Editor's Draft 4 May 2010

Latest Published Version:
http://www.w3.org/TR/html5/

Editors:

John Resig, Google Inc.

This specification defines the 5th major revision of the HTML (Hypertext Markup Language) in this version, new HTML elements are introduced, based on research into the related to defining clear conformance criteria for use of HTML.

If you wish to make comments regarding this document, please email html@w3.org.

Status of this document

This section describes the status of this document and lists any issues current to the working group. It also gives the latest date this document is valid. The status may be "Working Draft", "Editors' Draft", "Last Call" or "Published". A list of current W3C publications can be found at the W3C web site.
Standards support

Date and time

<time datetime="2004-08-08">8 สิงหาคม ๒๕๔๗</time>

Datetime picker demo

<form>
  <input type="date">
</form>
Standards support
Bidirectional text support

نشاط التدوير،

W3C

✔

✘
Developing requirements

Augmenting bidi support in HTML5 & CSS

Additional Requirements for Bidi in HTML

W3C Working Draft 4 March 2010

This version: http://www.w3.org/TR/2010/WD-html-bidi-20100304/

Latest version: http://www.w3.org/TR/html-bidi/

Editor: Aharon Lainin, Google

Additional Contributors:
- Adli Allawi, Technical Director, Diwan Software
- Mahtoua Alouche, Bidi Architect, IBM
- Uri Bernsstein, Google
- Douglas Davidson, Apple
- Mark Davis, Senior 118n Architect, Google; P118n
- Martin J. Dürst, W3C 118n Interest Group Chair
- Asmus Freytag, President, ASMU, Inc.
- Richard Ishida, 118n Lead, W3C
- Shanjian Li, Google
- Mohamed Mohie, IBM
- Jeremy Morkovich, Google
- Shachar Shemesh, Lingua Open Source Consulting
- Gaal Yahas, Google


Abstract

Authoring a web app that needs to support both right-to-left and left-to-right interfaces, or to take as input and display both left-to-right and right-to-left data, usually presents a number of challenges that make it an especially laborious and bug-prone task. Some of these are due to browser bugs, but some can be traced to a gap in the specification of...
Developing requirements

Arabic mathematics

Arabic mathematical notation

W3C Interest Group Note 31 January 2006

This version:
http://www.w3.org/TR/2006/NOTE-arabic-math-20060131

Latest version:
http://www.w3.org/TR/arabic-math/

Previous version:
This is the first version

Editors:
Azzeddine Lazrek, with Mustapha Eddahbi and Khadijatou翻身
Morocco
Bruce R. Miller, National Institute of Standards and Technology

This document is also available in these non-normative formats:

Copyright © 2006 W3C® (MIT, ERCIM, Keio). All Rights Reserved. W3C
Trustmark. W3C@Arabic Math 2006-01-31

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 in the MathML specification that require extensions in future versions.

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 document can 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 MathML and in the handling of Arabic mathematical presentation using MathML. It recommends some extensions and suggests extensions for a future revision.

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

Publication as an Interest Group Note does not imply endorsement by the W3C. It 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.
Press the <uitext translate="no">START</uitext> button to sound the horn. The <uitext translate="no">MAKE-READY/RUN</uitext> indicator flashes.

- supported by some translation tools – linked with XLIFF
- being applied by specifications at W3C
Internationalization & Localization metadata

<p translate="no">MultilingualWeb-LT</p>
The changing social context
The rise of the Mobile Web

Social context

- "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

Working with Time Zones

W3C Working Group Note 13 October 2005

Abstract
This document discusses some of the problems encountered when working with the date, time, and
datetime values from [XML Schema] when those values include (or omit) time zone offsets. Many W3C
technologies rely on date and time types. Examples include the [XPath1.0] specification, since it is the basis for
XQuery and XSLT processing of date/time values, but the concepts affect any date/time processing.

Status of this Document
This document is a W3C Working Group Note. It has been produced by the 11th Core Working Group, which is part of the
Internationalization Activity.

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

Tests

Internationalization (i18n) Activity

Text direction

This page groups together pages being developed by the HTML+CSS Internationalization Working Group to assess internationalization features still under development and should not be treated as a standard. Note that Internationalization WG tests do not specify standards. In some cases the tests also aim to test agents in ways not described by the standard.

Bidi algorithm in HTML

Basic block markup tests

Direction to element

1. paragraph
   HTML 4 HTML 5 XHTML 0.9
2. table
   HTML 4 HTML 5 XHTM 1.0

Link to tests: WOFF fonts

Characters should be right-to-left

i18n activity

 Assertions: In a language context, if an element containing mixed direcitons that are not in the same direction will not be in the same direction as the text.

i18n activity

<table>
<thead>
<tr>
<th>Font</th>
<th>WebKit</th>
<th>Firefox</th>
<th>Chrome</th>
<th>Opera</th>
<th>Safari</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Assertions: In a language context, if an element containing mixed direcitons that are not in the same direction will not be in the same direction as the text.
I18n resources

Best practices

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

I18n resources

Articles

Tutorials

Technical notes

Tests

Tools

Reviews

Internationalization (I18n) Activity
Making the World Wide Web usable

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
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
Best practices
I18n resources

Creating HTML & CSS
I18n resources

Best practices

Language

See also

The topic index organizes links to resources by keywords, rather than tasks (like the index in a book).

The resources by type page lists resources by type (e.g., articles, tool, mail archives, etc).

Current status

This index is still a work in progress. It doesn't yet point to all resources on the site. The content will also continually grow and change as resources are added to the site.
Choosing language values

- Identifying in-document language changes
- Indicating the language of a link destination
- Styling by language
- Using Accept-Language for locale setting
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
I18n resources

Best practices

Choosing a Language Tag

on this page: question - background - answer - by the way - further reading

Intended audience: XMTL, HTML coders (using editors or scripting), script developers (PHP, JSP, etc.), CSS coders, schema developers (DTDs, XML Schema, RelaxNG, etc.), XSLT developers, Web project managers, and anyone who needs guidance on how to construct language tags.

question

Which language tag is right for me? How do I choose language and other subtags?

[Skip to the answer]

background

In HTML and XML documents a language tag is used to indicate the language of content.

A language tag is composed of one or more subtags separated by hyphens. Subtags can be of various types.

Language tag syntax is defined by the IETF's BCP 47. In the past it was necessary to consult lists of codes in various ISO standards to find the right subtags, but now you only need to look in the IANA Language Subtag Registry. We will describe the new registry below.

This article provides advice on how to choose the components of a language tag. For an overview of the concepts defined in BCP 47, see Language tags in HTML and XML.

Addison Phillips and Mark Davis, authors of BCP 47, provided guidance during the writing of this article.

answer

Accessing the subtag registry

All the subtags you will need to create a language tag are found in one place, the IANA Language Subtag Registry. The registry is a long list file, containing nearly 6,000 entries.

The first (and often only) subtag in a language tag always designates a language. It is referred to in BCP 47 as the primary language subtag. We will use that term in this document to refer to the subtag that represents a language, to more clearly make the distinction from 'language tag,' which refers to the whole thing.

The notes on this page provide guidance that is sufficient for most people wanting to use language tags. There are links to detailed sections of BCP 47 in this region for people who want to read the full text of the specification.

Note: please that same environments or systems may display choices that are different from what you would otherwise expect. For example, in Spanish you must use "la" (depicted in BCP 47 in place of "el") (recommended in BCP 47).
<table>
<thead>
<tr>
<th>Language</th>
<th>Term</th>
<th>Views/Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
<td>views</td>
<td>0.8</td>
</tr>
<tr>
<td>English</td>
<td>次檢視</td>
<td>1.2</td>
</tr>
<tr>
<td>Portuguese</td>
<td>visualizações</td>
<td>2.6</td>
</tr>
<tr>
<td>French</td>
<td>consultations</td>
<td>2.6</td>
</tr>
<tr>
<td>German</td>
<td>angesehen</td>
<td>2.8</td>
</tr>
<tr>
<td>Italian</td>
<td>visualizazioni</td>
<td>3</td>
</tr>
</tbody>
</table>
Best practices

Text expansion

**Global settings**
- Interface language: English
- Search language: English
- Number of results: 10
- Save preferences

**Acuan Umum**
- Bahasa Pengantar di Antar Muka: Ingriss
- Bahasa Pengantar untuk Penelusuran: Ingriss
- Jumlah Hasil Penelusuran: 10
- Simpan Acuan

**Allgemeine Voreinstellungen**
- Sprache der Benutzeroberfläche: Englisch
- Suchsprache: Englisch
- Anzahl der Ergebnisse: 10
- Einstellungen speichern
Best practices

Checker tool

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

1. Discover
2. Check
Best practices
MultilingualWeb workshops

Developers
Creators
Localizers
Machines
Users
Policy

http://multilingualweb.eu
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.

• Review or contribute to development/dissemination of outreach materials, to help others understand how to implement and use international features of the Web.

• Attend the MultilingualWeb workshop in Luxembourg in March 2012.
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/1111-aslib/