वर्ल्ड वाइड वेब को सचमुच विश्वव्यापी बना रहें हैं!

Making the World Wide Web worldwide!

Richard Ishida
W3C Internationalization Activity Lead
Internationalization Activity

About the Consortium

• Help W3C Working Groups understand international issue and requirements as early as possible

• Check specifications in Working Draft stage, especially at Last Call, for internationalization issues

• Liaise with other standards organizations to develop support for the international Web

• Evangelize the need to consider multiple languages and scripts when developing Web technologies of any kind

• Help users of Web technology understand what's available to them and how to use it
Developing requirements
Developing requirements
International typography
Developing requirements
International typography

Additional Requirements for Bidi in HTML

W3C Working Draft 4 March 2010

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
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 Eddieh and Kees van Ginneken.
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 Recommendations are subject to change. Authors,ståndards, and software implementations may differ. Please refer to the W3C legal notice and acknowledgments, available at, http://www.w3.org/Conventions/notice-acknowledgements.html

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 may require extensions in future versions.

Status of this Document

This section describes the status of this document at the time of its publication. The content of this document may be superseded by other documents. A list of current W3C publications and the latest revised versions of all W3C technical reports 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 includes a detailed analysis of the handling of Arabic mathematical presentation using MathML. 2 Recommendations and suggestions for future extension are provided.

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

Publication as an Interest Group Note does not imply endorsement by the W3C.

The following symbols may be used:

\[ f(x) = \begin{cases} \sum_{i=1}^{s} x^i & \text{if } x < 0 \\ \int_{1}^{s} x^i \, dx & \text{if } x \in S \\ \tan \pi & \text{otherwise (with } \pi \approx 3.141) \end{cases} \]
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-synthesis-11-20100223/

Latest version:
http://www.w3.org/TR/speech-synthesis-11/

Previous version:

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

Authors:
Paolo Baglia, Logisero
Paul Bagshaw, France Telecom
Michael Bodel, Microsoft
胡方 (Hu Fang), Invited Expert (until 20 October 2009)
黄俊雄 (Zhi Huang), France Telecom
黄力行 (Li Xing Huang), Chinese Academy of Sciences
唐永国 (Yongguo Kang), Panasonic Corporation (until 5 December 2007)
柳锡煌 (Lou Xiaoyan), Toshiba
Scott McGlashan, HP
戴美玲 (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 90 October 2008)
夏浩明 (Xia Hao Ming), Panasonic Corporation (until 2 August 2008)
吴志勇 (Zhiyong Wu), Chinese University of Hong Kong (until 29 July 2009)
严骏 (Yan Jun), FLYTEK

Copyright © 2010 W3C®. All Rights Reserved. W3C liability, trademark and document usage rules apply.

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.
Speech Synthesis Markup Language (SML) Version 1.1

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:
http://www.w3.org/TR/2009/PR-speech-synthesis11-20090627/

Editors:
Daniel C. Burnett, Voxeo (formerly of Vocalocity and Nuance)
李林(Weihua Li), IBM

Authors:
Pablo Baggio, Loquendo
Paul Bagshaw, France Telecom
Michael Bodell, Microsoft
韩方(Shao Liang), Invited Expert (until 20 October 2009)
袁伟进(De Zhi Huang), Invited Expert (until 20 October 2009)
史强, (Lijun Huang), Chinese Academy of Sciences
藤倉田(Yonggao Kang), Panasonic Corporation (until 5 December 2007)
Intel (Liu Xiaoyan), Toshiba
Scott McGlashan, HP
蔡美玲(Helen Meng), Chinese University of Hong Kong (until 29 July 2009)
刘军华(Jianhua Tao), Chinese Academy of Sciences
王群 (Wang Xia), Nokia (until 30 October 2008)
寺阪保子(Kazuko Hara), Panasonic Corporation (until 2 August 2008)
吴志勇(Zhiyong Wu), Chinese University of Hong Kong (until 29 July 2009)
严俊 (Yan Jun), iFLYTEK

Copyright © 2010 W3C®. All Rights Reserved. W3C liability, trademark and document use rules apply.

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.
Integrating international features into specifications
Implementers of user agents need to be prodded by the public to support the developing marketplace!
<meta http-equiv="Content-Language" content="hi">

<link charset="utf-8" ...>

<input type="datetime" ...>
Internationalization Quicktips

- Use Unicode wherever possible for content, databases, etc. Always declare the encoding of content.

- Use characters rather than escapes (e.g. `&#xE1;`, `&#225;`; or `&aacute;`) whenever you can.

- Declare the language of documents and indicate in their title.
Coordinating with other standards
"The Path W3C follows to making text on the Web truly global is Unicode."

Tim Berners-Lee
<h2><a id="რჩეული">რჩეული</a></h2>

<p><a href="/wiki/ჭიამაია" title="ჭიამაია" class="mw-redirect">ჭიამაია</a> (Coccinellidae), ხოჭოების ოჯახს ეკუთვნის. აქვს ამობურცული, მომრგვალო ან ოვალური სხეული. ზურგზე ღია ფონზე შავი ლაქები აყრია, იშვიათად ...
რჩეული ფოტოსურათი
ჭიამაია (Coccinellidae), ხოჭოების ოჯახს ეკუთვნის. აქვს ამობურცულ, მომრგვალო ან ოვალური სხეული. ზურგზე ღია ფონზე შავი ლაქები აყრია, იშვიათად...
Coordination with other bodies

Normalization

NFD

Ízeliőtoőül

NFC

Ízelítőül

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 többnyelvű számítógépeken.
Coordination with other bodies

Web resource identifiers


Scheme

Domain name

Path

IDN

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

xn--jp-cd2fp15c.xn--fsq.jp
Coordination with other bodies

Web resource identifiers

السعودية.
امارات.
مصر.

Al-Saudiah
Emarat
Misr

http://وزارة-الاتصالات.مصر
Coordination with other bodies

Web resource identifiers


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

IRI: /dir1/%E5%BC%95%E3%81%8D%E5%89%B2%E3%82%8A.html
Coordination with other bodies

Language tags

- ISO 639 language codes
- ISO 3166 country codes
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
Education & outreach
Guidelines & education
Capturing guidance for spec developers

Working with Time Zones

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 revisions of this technical report can be found in the W3C technical reports index at http://www.w3.org/TR/TR/.

This document discusses the topic of date, time, and date/time values from [XML Schema] with and without time zone offsets. Examples are given mainly relying on [XML Schema] and [XPath 1.0], since these are the basis for [XQuery] and [XSLT 2.0] processing of date/time values.

This document is a W3C Working Group Note. It has been produced by the Internationalization Activity.
Tests

Guidelines & education
I18n resources

Guidelines & education

Articles

Tutorials

Technical notes

Tests

Talks

Tools

Reviews

http://www.w3.org/International/
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
Guidelines & education

I18n resources

W3C I18n technique index

Select the task you want help with from this page to find resources on the W3C Internationalization site among its articles, tutorials, tests and techniques documents.

General task

Select a general task...

- Using the Web
- Creating HTML & CSS
- Authoring SVG
- Authoring XML
- Setting up a server
- Developing specifications
- Developing schemas

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.
Guidelines & education

I18n resources

W3C I18n technique index

Select the task you want help with from this page to find resources on the W3C Internationalization site among its articles, tutorials, tests and techniques documents.

General task

Select a general task...
- Using the Web
- Creating HTML & CSS
- Authoring SVG
- Authoring XML
- Setting up a server
- Developing specifications
- Developing schemas

Creating HTML and CSS

Select a topic area...
- Characters
- Language
- Markup & text
- Text direction
- Styling & layout
- Forms
- Navigation
- Cultural issues
- Troubleshooting

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, tools, 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.

Author: Richard Ishida, W3C.
Content last changed 2009-03-12 16:23 GMT

Copyright © 2009 W3C (MIT, ERCIM, Keio). All Rights Reserved. W3C liability, trademark, document use and software licensing rules apply. Your interactions with this site are in accordance with our public and Member privacy statements.
W3C I18n technique index

Select the task you want help with from this page to find resources on the W3C Internationalization site among its articles, tutorials, tests and techniques documents.

General task      Creating HTML and CSS   Language

Select a general task...
- Using the Web
- Creating HTML & CSS
- Authoring SVG
- Authoring XML
- Setting up a server
- Developing specifications
- Developing schemas

Select a topic area...
- Characters
- Language
- Markup & text
- Text direction
- Styling & layout
- Forms
- Navigation
- Cultural issues
- Troubleshooting

Select a link...
- Getting started
- Using attributes to declare language
- Declaring metadata about the language of the intended audience
- Declaring language for documents aimed at speakers of more than one language
- Choosing language values
- Identifying in-document language changes
- Indicating the language of a link destination
- Styling by language
- Using Accept-Language for locale setting

Author: Richard Ishida, W3C.
Content last changed 2009-03-12 16:23 GMT

Copyright © 2009 W3C® (MIT, ERCIM, Keio), All Rights Reserved. W3C liability, trademark, document use and software licensing rules apply. Your interactions with this site are in accordance with our public and Member privacy statements.
Guidelines & education

I18n resources

Choosing language values

Recommendations:
- Follow the guidelines in the IETF’s BCP 47 for language attribute values.
- Use the shortest possible language tag values.
- Where possible, use the codes zh-Hans and zh-Hant to refer to Simplified and Traditional Chinese, respectively.
- Use the subtag xx when the text is known to be not in any language.
- If the XML format you are using supports it, use xml:lang="", otherwise use xml:lang="und" when the language is undetermined and you have to label it.

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 Content.
- Specifying language tag values
  In W3C tutorial, Declaring Language in XHTML and HTML.
- Tagging text with no language
  How do I use language markup in HTML or XML content when I don’t know the language, or the content is non-linguistic? W3C article.
- Two-letter or three-letter language codes
  Should I use two-letter or three-letter ISO language codes in language tags? W3C article.

Particularly useful links:
- IANA Language Subtag Registry
  This is the official location where you will find all subtags available for use in language tags.
- Language Subtag Lookup tool
  User friendly interface to IANA’s language tag registry by Richard Ishida.
  Provides for checking of subtags as well as lookup. Up-to-date with latest version of BCP 47.
- Internet-Draft: BCP 47
  Points to a document containing both RFC 5646 (Tags for the Identification of Languages) and RFC 4847 (Matching Language Tags).
- RFC 5646 Tags for the Identification of Languages
  The specification that describes language tag syntax.
- RFC 4647 Matching of Language Tags
  The specification that describes alternative ways of matching language tags.
Choosing a Language Tag

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

Intended audience: HTML, XML, CSS, and JavaScript coders (using editors or scripting), script developers (PHP, JSP, etc.), CSS coders, schema developers (RDFa, XML Schema, RelaxNG, etc.), XML transform 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 text file, containing nearly 9,000 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.
### Text expansion

<table>
<thead>
<tr>
<th>Language</th>
<th>Translation</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
<td></td>
<td>0.8</td>
</tr>
<tr>
<td>English</td>
<td>views</td>
<td>1</td>
</tr>
<tr>
<td>Chinese</td>
<td>次检视</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>-mal angesehen</td>
<td>2.8</td>
</tr>
<tr>
<td>Italian</td>
<td>visualizzazioni</td>
<td>3</td>
</tr>
</tbody>
</table>

*DochuLa Pass*

Around 3,100 m high, in Bhutan.

Uploaded on Jan 28, 2007 | Map | Delete
1,414 views / 5 comments
Guidelines & education

Text expansion

Seals and sealions at Pier 39
Guidelines & education

Text expansion

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

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

**Allgemeine Voreinstellungen**
- Sprache der Benutzeroberfläche: Englisch
- Suchsprache: Englisch
- Anzahl der Ergebnisse: 10

Save preferences
Simpan Acuan
Einstellungen speichern
### Checker tool

**W3C Internationalization Checker (Prototype only!)**

Is your Web site Internationalized?

<table>
<thead>
<tr>
<th>Address</th>
<th><a href="http://rishida.net/tools/i18nchecker/test.php">http://rishida.net/tools/i18nchecker/test.php</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Run another check</td>
<td></td>
</tr>
</tbody>
</table>

#### Information

<table>
<thead>
<tr>
<th>Character encoding</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTP Content-Type</td>
<td>No charset found.</td>
</tr>
<tr>
<td>Byte order mark (BOM)</td>
<td>UTF8</td>
</tr>
<tr>
<td>xml declaration</td>
<td>None found.</td>
</tr>
<tr>
<td>meta charset element</td>
<td>iso-8859-1</td>
</tr>
<tr>
<td>HTML5 meta charset element</td>
<td>None found.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;html lang=kk &gt;</td>
<td>&lt;html lang=&quot;kk&quot; xmlns=&quot;<a href="http://www.w3.org/1999/xhtml">http://www.w3.org/1999/xhtml</a>&quot; &gt;</td>
</tr>
<tr>
<td>&lt;html xml:lang=ko &gt;</td>
<td>&lt;html lang=&quot;kk&quot; xml:lang=&quot;ko&quot; dir=&quot;ltr&quot; xmlns=&quot;<a href="http://www.w3.org/1999/xhtml">http://www.w3.org/1999/xhtml</a>&quot; &gt;</td>
</tr>
<tr>
<td>HTTP Content-Language</td>
<td>ka, ta</td>
</tr>
<tr>
<td>meta content-language element</td>
<td>en,fr,sp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Text direction</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default direction</td>
<td>ltr</td>
</tr>
<tr>
<td>&lt;html lang=kk &gt;</td>
<td>&lt;html lang=&quot;kk&quot; dir=&quot;ltr&quot; xmlns=&quot;<a href="http://www.w3.org/1999/xhtml">http://www.w3.org/1999/xhtml</a>&quot; &gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class &amp; id names</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-ascii class or id names</td>
<td>8</td>
</tr>
<tr>
<td>Non-NFC class or id names</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Request headers</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept-Language</td>
<td>en,fr,sp;q=0.8,fr;q=0.6,ch;q=0.3</td>
</tr>
<tr>
<td>Accept-Charset</td>
<td>ISO-8859-1,utf-8;q=0.7,*;q=0.7</td>
</tr>
</tbody>
</table>
Detailed report

Severity Description

- Multiple encoding declarations using the meta tag.
  
a. `<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />`
b. `<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />`

One document has to have a single character encoding. You should edit the markup to remove all but one declaration.

- Incorrect values used for dir attribute.
  
  a. `<p dir="il" lang="he" xml:lang="he" title="Hebrew : Hebrew" class="phrase">`  
   The value should be one of 'rtl' or 'ltr'.

- Class or id names found that are not in Unicode Normalization Form C.
  
  Unicode allows you to represent certain letters using different combinations of bytes. To avoid problems when trying to match class or id names against CSS selectors or JavaScript lookup, the text for all your markup tags and CSS and JavaScript code should use the same byte combinations for the same text, i.e. be normalised. We recommend Unicode Normalized Form C (NFC).
  
  Total number of non-NFC names: 4.
  
a. `<p title="Latin : Hungarian" lang="hu" xml:lang="hu" class="phrase Nemzetkozlosites">`
b. `<p title="Latin : Welsh" lang="cy" xml:lang="cy" class="phrase id="Terulet">`
c. `<div id="Gam">`
d. `<p id="san">`

For more information about working with NFC, see Unicode normalization forms in the tutorial Character encodings in HTML and CSS.

- Conflicting character encoding declarations.
  
The HTTP Content-Language information doesn't match the lang attribute value in the html tag.

- The lang attribute and the xml:lang attribute in the html tag have different values.

- A lang attribute value did not match an xml:lang value when they appeared together on the same tag.

- A language attribute value was incorrectly formed.

- A tag uses a lang attribute without an associated xml:lang attribute.
Moving forward...
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)

- Read and review W3C specifications and send comments to the i18n lists, and be prepared to follow up with 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.

- Take on board that internationalization is something done by developers and designers – not localizers. Find out how to do it. (http://www.org/International/)

- 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
- if something isn't right, get involved to fix it, and encourage local developers to do the same
- start by visiting http://www.w3.org/International/