The Multilingual Web: Latest developments at the W3C/IETF
Outline

- About the W3C
- Standards support for the multilingual Web
- The changing social context
- Best practices for the multilingual Web
- Getting involved
About the W3C
About the Consortium


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

Keywords:
consensus and vendor neutrality
About the Consortium

Accountable to the global public!

Technology & Society
- eGovernment
- Patent Policy
- Privacy
- Security

Technology development
- Fonts
- Graphics (SVG, WebCGM)
- HTML
- Math
- Mobile Web
- Multimodal Interaction
- Rich Web Client (WebApps)
- Semantic Web
- Style (CSS)
- Synchronized Multimedia
- Ubiquitous Web Apps
- Video in the Web
- Voice Browser
- Web Services
- Xforms
- XML

Web for All
- Internationalization
- WAI International Program
- Office
- WAI Technical Activity

22 Activities
50 Working Groups
13 Interest Groups
5 Coordination Groups
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
"The Path W3C follows to making text on the Web truly global is Unicode."

Tim Berners-Lee
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/2006/PER-xml-20060206/
http://www.w3.org/TR/2006/REC-xml-20060616/

Editors:
Tim Bray, Textuality and Netscape <tbray@textuality.com>
Jean Paoli, Microsoft <jeanpa@microsoft.com>
C. M. Sparber McQueen, W3C <cmqm@w3.org>
Eve Maler, Sun Microsystems, Inc. <eve.maler@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 Web. It extends the SGML syntax for the purpose of supporting implementation and interoperability. Computer programs that process XML documents are expected to be SGML-aware, in the sense that the SGML keywords and their respective SGML tags are recognized.

XML is a language that enables the exchange of data in a non-proprietary, generic format on the Internet. XML is defined by a simple declarative language that describes the syntax and structure of a document. XML documents can be structured in a wide variety of ways, such as using DTDs, XML Schemas, or other XML vocabulary. XML documents can be transformed into HTML, plain text, or any other format using XSLT.

XML documents are not intended to be processed by human beings, but are meant to be processed by computer programs that support XML. XML documents can be validated against an XML Schema or DTD, which describes the structure and content of the XML document. XML documents can also be transformed into other formats using XSLT, which specifies how the XML document is to be transformed into another format.

XML documents are intended to be useful for a wide variety of applications, such as exchanging data between different systems, or exchanging data with a human-readable format. XML documents are intended to be self-contained, so that they can be processed independently of other XML documents. XML documents are also intended to be reusable, so that they can be used as building blocks for other XML documents.

XML documents are intended to be portable, so that they can be processed on any platform. XML documents are intended to be scalable, so that they can be processed on any computer system. XML documents are also intended to be extensible, so that they can be extended to support new applications.
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, szövegei rendezésnél, és többnyelvű számítógépeken.
Web resource identifiers


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

IDN: xn--jp-cd2fp15c.xn--fsq.jp
Standards support

Web resource identifiers

السعودية
امارات
مصر

Al-Saudiah
Emarat
Misr

http://وزارة-الاتصالات.مصر

IDN
Web resource identifiers

 Standards support

 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
Language tags

- ISO 639 language codes
- ISO 3166 country codes

Before: RFC 3066

language – region

en
en-GB
en-scouse
Standards support

Language tags: BCP 47

Now: BCP 47 (includes RFC 5646)

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

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

- Use characters rather than escapes (e.g. \xE1; \x225; or &aacute;) 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-synthesis11-20100222/

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

Previous version:

Editors:
- Paolo Bagnia, Logonodo
- Paul Bagshaw, France Telecom
- Michael Bodell, Microsoft
- 钟方 (Hu Fang), Invited Expert (until 20 October 2009)
- 黄俊华 (Da Zhi Huang), France Telecom
- 黄力行 (Ji-Xin Huang), Chinese Academy of Sciences
- 康永军 (Yongjun Kang), Panasonic Corporation (until 5 December 2007)
- 嵇晓勇 (Lou Xiaoyan), Toshiba
- Scott McGlasson, HP
- 茹笑玲 (Helen Meng), Chinese University of Hong Kong (until 29 July 2009)
- 隋езн (jianhua Tao), Chinese Academy of Sciences
- 王霞 (Wang Xiao), Nokia (until 30 October 2008)
- 王蒋竞 (Xia Haining), Panasonic Corporation (until 2 August 2008)
- 吴志勇 (Zhiyong Wu), Chinese University of Hong Kong (until 29 July 2009)
- 严俊 (Yan Jun), FLYTEK

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.
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.
Standards support

CSS3

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

Vertical text

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
網際網路及統一碼發展、
國際化及本土化、支援統
一碼的作業系統及應用程
式、字型、文字排版、電
腦多國語文化等多項課
題。
Standards support

Vertical text
Standards support
Ruby annotation
Standards support

Ruby annotation

凝ぎょう視し

凝ぎょう視し

凝ぎょう視し

<ruby><rb>凝</rb><rb>視</rb></ruby>ぎょうし

<ruby><rb>凝</rb><rb>視</rb></ruby>ぎょうし

<ruby><rb>凝</rb><rb>視</rb></ruby>ぎょうし
Requirements for Japanese Text Layout

Development 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 (岡田 雅宏), Microsoft
Hiroaki Chiba (千葉裕章), Invited Expert
Junsaburo Edamoto (枝本順三郎), Invited Expert
Richard Ishida, W3C
Keiichiro Ishino (石野 喜一郎), Antenna House
Tatsuo Kobayashi (小林光生), JustSystems
Toshi Kobayashi (小林敏), Invited Expert
Kenzou Onozawa (小野澤 眞三), 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.

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.
Standards support
Language declarations

<html lang="de">
<head>

<meta http-equiv="Content-Language" content="de">

</head>

</html>
Standards support

Date and time

<time datetime="2004-08-08">๘ สิงหาคม ๒๕๔๗</time>

Datetime picker demo

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

 Neuroscience, W3C

ackle التدويل، W3C

ackle التدويل، W3C

<description dir="rtl">W3C نشاط التدويل،</description>
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 Lain, Google

Additional Contributors:
Adil Allawi, Technical Director, Diwan Software
Mattiafu Alouche, Bidi Architect, IBM
Uri Bernstein, Google
Douglas Davidson, Apple
Mark Davis, Senior 11th Architect, Google; P
Martin J. Durst, W3C 11th Interest Group Chair
Asmus Freytag, President, ASMUS, Inc.
Richard Ishida, 11th Lead, W3C
Shanjian Li, Google
Mohamed Mohie, IBM
Jeremy Moskovich, Google
Shachar Shemesh, Lingui 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 2008

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 Note 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 2.1. This Note recommends the handling of Arabic mathematical presentation using MathML 2.1 Recommendations and suggests extensions for a future revision.

This Note has been written by participants in the Math Interest Group (W3C member group) of the W3C Math activity. Please direct comments and report errors in this document to the Math Interest Group's public archive.

Publication as a Math Interest Group Note does not imply endorsement by the W3C. This document 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.

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 role of this document 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.

\[
\begin{align*}
\sum_{i=1}^{s} x^i & \quad \text{if } x < 0 \\
\int_{1}^{s} x^i \, d\! x & \quad \text{if } x \in S \\
tan \pi & \quad \text{otherwise (with } \pi \approx 3.141) \\
\end{align*}
\]

ت (س) = \begin{cases} 
\frac{ب}{ب} & \text{إذا كان } س > 0 \\
\frac{ب}{ب} & \text{إذا كان } س \leq M \\
\pi & \text{غير ذلك (مع } \pi \text{)} 
\end{cases}
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
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

Character Model

W3C Recommendation

Abstract

This Architectural Specification defines a Character Set, defined jointly with
the terms 'character', 'coding', 'character encoding', 'character encoding
for normalization and string matching', and 'Normalization Model for the
World Wide Web 1.0' Normalization Model for the World Wide Web

Status of this Document

This section describes the status of this document. The current version can be
found in the W3C technical report index at http://www.w3.org/TR/.

Status of the 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/.

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 [XPath 2.0] specification, since it is the basis for
XQuery and XSLT processing of datetime values, but the concepts apply to any date / time processing.

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

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

Tests
I18n resources

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
I18n resources

Creating HTML & CSS

W3C I18n technique index

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

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.
Best practices

I18n resources

Language
Best practices

I18n resources

How to's

W3 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.

W3 Language tags in HTML and XML
A simple overview of the syntax for language tags in BCP 47. W3C article.

W3 How to choose language values
In W3C techniques document, Specifying Language in XHTML and HTML
Choosing a Language Tag

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

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 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.
### Text expansion

<table>
<thead>
<tr>
<th>Language</th>
<th>Translation</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
<td>次檢視</td>
<td>0.8</td>
</tr>
<tr>
<td>English</td>
<td>views</td>
<td>1</td>
</tr>
<tr>
<td>Chinese</td>
<td>次視</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>-mal angesehen</td>
<td>2.8</td>
</tr>
<tr>
<td>Italian</td>
<td>visualizzazioni</td>
<td>3</td>
</tr>
</tbody>
</table>
Best practices

Text expansion
Best practices
Checker tool

1. Discover
2. Check

http://qa-dev.w3.org/i18n-checker/
The social context

Mobile Web Initiative

Making Web access from a mobile device as simple as Web access from a desktop device.

The Mobile Web Initiative's goal is to make browsing the Web from mobile devices a reality, explains Tim Berners-Lee, W3C Director and inventor of the Web. "W3C and mobile industry leaders are working together to improve Web content production and access for mobile users and the greater Web."

Mobile Web Initiative participants are developing best practices for creating mobile-friendly content and applications, enabling easy access to device descriptions, setting up test suites for increased interoperability of mobile browsers, and exploring ways to use the Web on mobile devices to bridge the digital divide.

NEXT MWI EVENT

Most of the MWI staff will be present at the ICT 2008 event, organized by the European Commission, on 25-27 Nov. 08, in Lyon, France. Marie-Claire, Philipp, and Stéphane will be at booth 606, in the International Village.

CHECK YOUR SITE / YOUR BROWSER

Is your site mobileOK?

Drag the mobileOK bookmarklet to your toolbar to quickly check pages: mobileOK check

How does your mobile browser fare in the Web Compatibility Test?
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/