

वर्ल्ड वाईड वेब को सचमुच  
विश्वव्यापी बना रहें हैं !

Making the World Wide Web worldwide!

Richard Ishida

W3C Internationalization Activity Lead



# Internationalization Activity

- 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

عالمگیر ویب کو تحقیقی طور پر عالمگیر بنانا

ՀՀԼ ΔΡԿԹՃԻ ԴԸԿՎԸԼԴԻ օՆՈՂՆՆՆ

"Дүниежүзілік торды" нағыз дүниежүзілік етеміз!  
बल्ड 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 ਬਣਾਉਣਾ !

缔造真正全球通行的万维网

ワールド・ワイド・ウェブを世界中に広げましょう  
ធ្វើឲ្យវីលវ៉ាយវ៉ែបមានទូទាំងពិភពលោកពិប្រាកដមែន!

전세계의 월드 와이드 웹으로 만들기!

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

ליצור מהרשת הרשת לל עולמית באמת!

การทำให้ World Wide Web แพร่หลายไปทั่ว

โลกอย่างแท้จริง

འཛམ་གླིང་ཡོངས་འབྲེལ་འདི་ རོ་མ་འབད་རང་ འཛམ་གླིང་ཡོངས་ལུ་ཁུ་བཟུགས་པ་  
བཟོ་བ།



# Developing requirements

# Developing requirements

## International typography



[\[ contents \]](#)

### Requirements for Japanese Text Layout

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 [translation](#).  
 English version of this document is the authoritative version.

Copyright © 2008-2009 W3C® ([MIT](#), [ERCIM](#), [Keio](#)). All Rights Reserved. W3C [liability](#), [trademark](#), [document use](#) rules apply.

### Abstract

This document describes requirements for general Japanese layout realized by modern 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.

The diagram shows various examples of Japanese text with annotations for spacing and alignment:

- の流儀を** (の りゆうぎを): Shows the character 'の' followed by a full-width stop and comma, then '流儀' and 'を'. A note indicates 'Character advance of full stops and commas is half-width. 1/2 em space after'.
- の無常を** (の むじょうを): Shows the character 'の' followed by a full-width stop and comma, then '無常' and 'を'. A note indicates 'Character advance of brackets is half-width. 1/2 em space before or after'.
- の成就を** (の じょうじゆを): Shows the character 'の' followed by a full-width stop and comma, then '成就' and 'を'. A note indicates 'Character advance of brackets is half-width. 1/2 em space before or after'.
- の紋** (の もん): Shows the character 'の' followed by a full-width stop and comma, then '紋'. A note indicates 'Character advance of middle dots is 1/4 em space'.
- の象** (の しょう): Shows the character 'の' followed by a full-width stop and comma, then '象'. A note indicates 'Character advance of middle dots is 1/4 em space'.
- その前に** (その 前に): Shows the characters 'その' followed by a full-width stop and comma, then '前'. A note indicates 'Character advance of middle dots is 1/4 em space'.
- 中心となる** (中心 となる): Shows the characters '中心' followed by a full-width stop and comma, then 'となる'. A note indicates 'Character advance of brackets is half-width. 1/2 em space before or after'.
- 本の上を天、下部を地、前側を** (本の上を天、下部を地、前側を): Shows the characters '本の上を天、下部を地、前側を'. A note indicates 'Character advance of brackets is half-width. 1/2 em space before or after'.
- 小口、その反対をのどという。** (小口、その反対をのどという。): Shows the characters '小口、その反対をのどという。'. A note indicates 'Character advance of brackets is half-width. 1/2 em space before or after'.
- 君子は和して同ぜず** (君子は和して同ぜず): Shows the characters '君子は和して同ぜず'. A note indicates 'Character advance of brackets is half-width. 1/2 em space before or after'.

くん し わ どう  
 君子は和して同ぜず

# Developing requirements

## International typography

[\[ contents \]](#)



### 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 Lanin, Google

**Additional Contributors:**

- Adil Alawi, Technical Director, Diwan Software
- Matitiahu Allouche, Bidi Architect, IBM
- Uri Bernstein, Google
- Douglas Davidson, Apple
- Mark Davis, Senior I18n Architect, Google; P
- Martin J. Dürst, W3C I18n Interest Group Chair
- Asmus Freytag, President, ASMUS, Inc.
- Richard Ishida, I18n Lead, W3C
- Shanjian Li, Google
- Mohamed Mohie, IBM
- Jeremy Moskovich, Google
- Shachar Shemesh, Lingnu Open Source Consulting
- Gaal Yahas, Google

Copyright © 2007-2010 W3C® (MIT, ERCIM, Keio). All Rights Reserved. W3C [liability](#), [trademark](#) and [document use](#) rules apply.

### 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

TOP RATED  
RESTAURANTS

Aroma - 3 reviews



TOP RATED  
RESTAURANTS

Aroma - 3 reviews



### Translate text, webpages and documents

Enter text or a webpage URL, or [upload a document](#).

Internationalization Activity, W3C

Translate from: English

Translate into: Arabic

### Translate text, webpages and documents

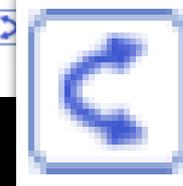
Enter text or a webpage URL, or [upload a document](#).

ضابطه لساني عدميت، W3C

Translate from: English

Translate into: Arabic

Translate



# 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 Eddahibi and Khalid  
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

### 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. It may be superseded by other W3C publications and the latest revision of this document. A list of current W3C publications and the latest revision of this document can be found in the [W3C technical reports index](http://www.w3.org/TR/) at <http://www.w3.org/TR/>.*

This Note is a self-contained discussion of Arabic mathematical notation in MathML 2 Recommendation and suggests extensions for a future revision.

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

Publication as a Working Group Note does not imply endorsement by the W3C. The 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.

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

$$\left. \begin{array}{l} \text{مجا } \frac{\text{ص}}{\text{س ب}} \\ \text{ب} = \text{ا} \\ \text{ا } \text{س ب} \text{ ء س} \\ \text{ظا } \pi \end{array} \right\} = (\text{س}) \text{ ت}$$

إذا كان س > 0  
إذا كان س ∈ م  
غير ذلك (مع π ≈ 3.141)

# Speech Synthesis Markup Language

W3C Proposed Recommendation



## 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:**

<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 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.

different synthesis-capable platforms.  
content a standard way to control aspects of speech such as pronunciation, volume, pitch, rate, etc. across  
Web and other applications. The essential role of the markup language is to provide authors of synthesizable

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



## Workshop on Conversational Applications

Speech Synthesis Markup Language (SSML) 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/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 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.

## Use Cases and Requirements for New Models of Human Language to Support Mobile Conversational Systems

June 18-19, NJ, USA

different synthesis-capable platforms.  
content a standard way to control aspects of speech such as pronunciation, volume, pitch, rate, etc. across  
Web and other applications. The essential role of the markup language is to provide authors of synthesizable



# Integrating international features into specifications



### CSS3 module: line

Editor's Draft 23 February 2008

**This version:**

[no permanent URL]

**Latest version:**

<http://www.w3.org/TR/css3-linebox>

**Previous version:**

<http://www.w3.org/TR/2002/WD-css3-linebox-20020515/>

**Editors:**

[Michel Suignard](#) (Microsoft) (until February 2003)

[Eric A. Meyer](#) (until February 2003)

[L. David Baron](#) (Mozilla Corporation) (from March 2003)

**Authors:**

Original CSS2 authors, and

[Bert Bos](#) (W3C)

[Michel Suignard](#) (Microsoft)

[Steve Zilles](#) (Adobe)

[L. David Baron](#) (Mozilla Corporation)

Copyright © 2008 W3C® (MIT, ERCIM, Keio). All Rights Reserved. W3C liability, trademark and document copyright notices.

### Abstract

The CSS formatting model provides for a flow of elements and text inside of a container to be wrapped into lines. The formatting of elements and text within a line, its positioning in the inline progression direction, and the breaking of lines are described in [CSS3TEXT]. This module describes the positioning in the block progression direction both of elements and text within lines and of the lines themselves. This positioning is often relative to a baseline. It also describes special features for formatting of first lines and drop caps. It extends on the model in [CSS2].

### Status of This Document

This is a public copy of the editors' draft. It is provided for discussion only and may change at any moment. Its publication here does not imply endorsement of its contents by W3C. Don't cite this document other than as work in progress.

The (archived) public mailing list [www-style@w3.org](mailto:www-style@w3.org) (see instructions) is preferred for discussion of this specification. When sending e-mail, please put the text "css3-linebox" in the subject, preferably like this: "[css3-linebox] summary of comment".

hanging

alphabetic

क्रिकेट के लिए मशहूर व इसके लिए दीवानगी दिखाने वाले खेल प्रेमियों के देश भारत में अगर कोई अन्य खिलाड़ी

स्थिति



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



Advising Working Groups

Web fonts

തൊണ്ണൂറ്  
ണ്ണൂറ്

உவலாஜயகி யகய  
ய

کو قفقی طور پر عالمگیر بنانا

ཀྱལ་སྡེར་བསྐྱར་བའི་ལས་དོན།

# Advising Working Groups

## HTML5

```
<meta  
  http-equiv="Content-Language"  
  content="hi">
```

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

```
< input type="datetime" ...
```

W3C Editor's Draft



### 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/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/>

**Editors:**  
[Jan Hickson](#), Google, Inc.

This specification is available in the following formats: [single page HTML](#), [multipage HTML](#). This is revision \$Revision: 1.4073 \$.

Copyright © 2010 W3C®. MIT, ISC, and DISA. All Rights Reserved. W3C liability, disclaimer, and document use rules apply.

---

### 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, new elements are introduced based on research into prevailing authoring practices, and special attention has been given to defining clear conformance criteria for user agents in an effort to improve interoperability.

### 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](http://www.w3.org/TR/) at <http://www.w3.org/TR/>.*

If you wish to make comments regarding this document, please send them to [public-html-comments@w3.org](mailto:public-html-comments@w3.org)

### Edit the photo date

at

< APR      MAY      JUN >      (date, and 24hr format for the time.)  
< 2006      2007      2008 >

SUN	MON	TUE	WED	THU	FRI	SAT
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		



Date Taken displays in the owner's timezone.

# Document Object Model Level 3 Key Events

W3C Editor's Draft



## Document Object Model (DOM) Level 3 Events Specification

W3C Editor's Draft 21 April 2010

**This version:**  
<http://dev.w3.org/2006/webapi/DOM-Level-3-Events.html?rev=1.131>

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

**Previous version:**  
<http://dev.w3.org/cvsweb/~checkout-/2006/webapi/html/DOM3-Events.html?rev=1.130>

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

**Editors:**  
 Doug Schepers, *W3C*  
 Björn Höhrmann, *Invited Expert (until December 2009)*  
 Philippe Le Hégarret, *W3C (until November 2003)*  
 Tom Pixley, *Netscape Communications Corporation*

Copyright © 2009 W3C® (MIT, ERCIM, Keio). All Rights Reserved. W3C and the W3C logo are registered trademarks and service marks of W3C. All other marks contained herein are the property of their respective owners. Use rules apply.

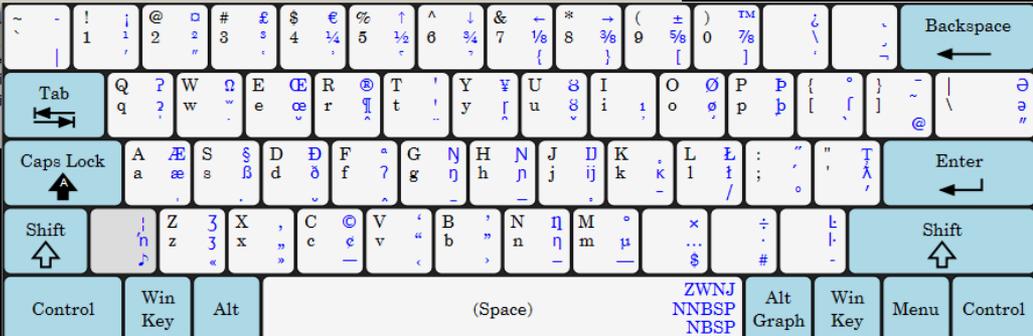
### Abstract

This specification defines the Document Object Model (DOM) Level 3 Events Specification, a platform- and language-neutral event system which allows registration of event listeners and describes event flow through a tree structure, and provides basic information for each event. The Document Object Model Events Level 3 builds on the Document Object Model Events Level 2 [DOM2 Events].

SEARCH MOBILE ACCESSIBILITY I18N TYPOGRAPHY

## 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 &acute;) whenever you can.
- ▶ Declare the language of documents and indicate in





# Coordinating with other standards



# Unicode

جعل شبكة الويب العالمية عالمية حقاً!

وب جهانی را درستی جهانی سازیم!

عالمگیر ویب کو تحقیقی طور پر عالمگیر بنانا

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

Համաշխարհային ցանցն իրոք համաշխարհային դարձնելը

ՀԵՂԼ ΔΡ<sup>α</sup>ΡΔ<sup>β</sup> Γ<sup>ε</sup>Δ<sup>α</sup>Γ<sup>β</sup>

"Дүниежүзілік торды" нағыз дүниежүзілік деміз!

वर्ल्ड वाईड वेबलाई यथार्थमे विश्वव्यापी बनाउने

የዓለም አቀፋን ድር በእውነት አለም አቀፍ

Κάνοντας τον Παγκόσμιο Ιστό πραγματικά Παγκόσμιο

ਵਰਡ ਵਾਈਡ ਵੈਬ ਨੂੰ ਵਾਕਈ ਵਿਸ਼ਵ-ਵਿਆਪੀ ਬਣਾਉਣਾ !

Tim Berners-Lee

締造真正全球通行的万维网

ליצור מהרשת רשת כלל עולמית באמת!

'meikɪŋ ðə wɜ:lɪd waɪd web 'tru:lɪ 'wɜ:lɪd'waɪd

ワールド・ワイド・ウェブを世界中に広げましょう

ធ្វើឲ្យវើបវ៉ាយវើបមានទូទាំងពិភពលោកពិប្រាកដមែន!

전세계의 월드 와이드 웹으로 만들기!

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

การทำให้ World Wide Web แพร่หลายไปทั่วโลกอย่างแท้จริง

འཛམ་གླིང་ཡོངས་འབྲེལ་འདི་ རོ་མ་འབད་པའ་ འཛམ་གླིང་ཡོངས་ལུ་བྱུང་ཚུགས་པ་བཟོ་བ།

# Unicode

```
<h2><a id="რჩეული" >რჩეული  
ფოტოსურათი</a></h1>
```

```
<p><a href="/wiki/ჭიამია" title="ჭიამია"  
class="mw-redirect">ჭიამია</a> (Coccinellidae),  
ხოჭოების ოჯახს ეკუთვნის. აქვს ამობურცული,  
მომრგვალო ან ოვალური სხეული. ზურგზე ღია  
ფონზე შავი ლაქები აყრია, იშვიათად
```

...

# Coordination with other bodies

## Unicode

<h2><a id="რჩეული">  
ფოტოსურათი</a></h2>

<p><a href="/wiki/ქიამა  
class="mw-redirect">ქიამა  
ხოქობის ოჯახს ეკუთვნის  
მომრგვალო ან ოვალური  
ფონზე შავი ლაქები აქვს  
...

W3C Recommendation



## 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](mailto:tbray@textuality.com)>

Jean Paoli, Microsoft <[jeanpa@microsoft.com](mailto:jeanpa@microsoft.com)>

C. M. Sperberg-McQueen, W3C <[cmsmcq@w3.org](mailto:cmsmcq@w3.org)>

Eve Maler, Sun Microsystems, Inc. <[eve.maler@east.sun.com](mailto: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 Web. This document specifies a syntax created by subsetting an existing, widely used international text processing standard

[technical reports index](#) at <http://www.w3.org/TR/>.

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/>.

### Status of this Document

## Normalization

NFD

l ő z e l i ő t o ő ű u ő l

NFC

Í z e l í t ő ű l

Ha a világ beszélni akarna, Unicode-ul szólalna meg. Regisztráljon már most a Tizedik Nemzetközi Unicode Konferenciára, melyet 1997. március 10-12-én rendeznek Mainz-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, szövegrendezésnél, és többnyelvű számítógépeken.



Coordination with other bodies  
Web resource identifiers

http://JP納豆.例.jp/dir1/引き割り.html

Scheme

Domain name

Path

IDN

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

Coordination with other bodies  
Web resource identifiers

السعودية.

Al-Saudiah

امارات.

Emarat

مصر.

Misr

<http://وزارة-الاتصالات.مصر>

IDN



Coordination with other bodies  
Web resource identifiers

http://JP納豆.例.jp/dir1/引き割り.html

Scheme

Domain name

Path

IRI

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



Coordination with other bodies

# Language tags

Before: RFC 3066

language – region

- ◆ ISO 639 language codes
- ◆ ISO 3166 country codes

en

en-GB

en-scouse



Coordination with other bodies

# Language tags: BCP 47

Now: BCP 47 (includes RFC 5646)

language    script    region    variant    extension    private\_use

hi

az-Cyrl

zh-Hans

es-419

sl-IT-rozaj-njiva-1994

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



# Education & outreach

# Capturing guidance for spec developers

W3C Recommendation



## Character Model

W3C Recommendation

**This version:**  
<http://www.w3.org/TR/>

**Latest version:**  
<http://www.w3.org/TR/>

**Previous version:**  
<http://www.w3.org/TR/>

**Editors:**  
 Martin J. Dürst, W3C  
 François Yergeau (Invited Expert)  
 Richard Ishida, W3C  
 Misha Wolf (until Dec 2005)  
 Tex Texin (Invited Expert)

Please refer to the [errata](#) for this document.

See also [translations](#).

Copyright © 2005 W3C® (MIT, ERCIM, Keio)

### Abstract

This Architectural Specification defines a common reference for the Character Set, defined jointly by the terms 'character', 'encoding', 'character encodings, character sets, and character models.

For normalization and string comparison, see *Wide Web 1.0: Normalization Model for the World Wide Web*.

### Status of this Document

This section describes the status of 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](#).

This document contains the [W3C Recommendation](#). It has been produced by the [W3C Working Group](#).

W3C Working Group Note



## Best Practices

W3C Working Group Note

**This version:**  
<http://www.w3.org/TR/>

**Latest version:**  
<http://www.w3.org/TR/>

**Previous version:**  
<http://www.w3.org/TR/>

**Editors:**  
 Yves Savourel, ENL  
 Jirka Kosek, Invited Expert  
 Richard Ishida, W3C

This document is also available in these non-normative formats: [XML](#).

Copyright © 2008 W3C® (MIT, ERCIM, Keio)

### Abstract

This document provides a common reference for the internationalized property, applications, as well as the character set, character encodings, character sets, and character models.

### Status of this Document

This section describes the status of 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](#).

This is a [W3C Working Group Note](#) developed by the [Internationalization Activity](#).

Feedback about this document is invited. Please send comments to [comment@w3.org](mailto:comment@w3.org). This list is publicly available.

Publication as a Working Group Note does not imply endorsement by the W3C as a whole.

[\[ contents \]](#)



## Working with Time Zones

W3C Working Group Note 13 October 2005

**This version:**  
<http://www.w3.org/TR/2005/NOTE-timezone-20051013/>

**Latest version:**  
<http://www.w3.org/TR/timezone>

**Editors:**  
 Addison Phillips, (Invited Expert)  
 Felix Sasaki, W3C  
 Mark Davis, IBM  
 Martin Dürst, Aoyama University

This document is also available in these non-normative formats: [XML](#).

Copyright © 2005 W3C® (MIT, ERCIM, Keio), All Rights Reserved. W3C [liability](#), [trademark](#) and [document use](#) rules apply.

### Abstract

This document discusses some of the problems encountered when working with the `date`, `time`, and `dateTime` values from [XML Schema](#) when those value include (or omit) time zone offsets. Many W3C technologies rely on date and time types. Examples include the [XPathFO](#) 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 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 [XPathFO](#), 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 [I18n Core Working Group](#), which is part of the [Internationalization Activity](#).

[\[ contents \]](#)

markup for  
bidirectional text

Normalization

working with case  
sensitivity

more information  
about date & time

**W3C Internationalization (I18n) Activity**  
*Making the World Wide Web world wide!*

Home Resources Techniques Topics News Groups About

## HTML+CSS Internationalization Text direction

This page groups together pages being developed by the Working Group to assess internationalization still under development and should not be used as a reference.

Note that Internationalization WG tests do not conform to standards. In some cases the tests also allow agents in ways not described by the standards.

### Bidi algorithm in HTML (no markup)

1. directional type, ltr context, hebrew  
HTML4 HTML5 XHTML1.0.html
2. directional type, ltr context, arabic  
HTML4 HTML5 XHTML1.0.html
3. directional type, ltr context, hebrew  
HTML4 HTML5 XHTML1.0.html
4. directional type, rtl context, english  
HTML4 HTML5 XHTML1.0.html
5. directional type, rtl context, english  
HTML4 HTML5 XHTML1.0.html
6. directional type, rtl context, english  
HTML4 HTML5 XHTML1.0.html
7. numbers, european  
HTML4 HTML5 XHTML1.0.html
8. numbers, arabic  
HTML4 HTML5 XHTML1.0.html

### Basic block markup tests [results]

#### Page direction set to RTL

1. paragraphs  
HTML4 HTML5 XHTML1.0.html
2. tables  
HTML4 HTML5 XHTML1.0.html

#### Direction on element

1. ltr context, rtl p  
HTML4 HTML5 XHTML1.0.html
2. ltr context, rtl table  
HTML4 HTML5 XHTML1.0.html
3. rtl context, ltr p  
HTML4 HTML5 XHTML1.0.html
4. rtl context, ltr table  
HTML4 HTML5 XHTML1.0.html

I18n test: ltr context, direction rtl, unicode-bidi

ltr context, direction rtl, unicode-bidi

Characters should be in LTR context

i18n activity "التدويل"

i18n activity "التدويل"

Assertion: In a LTR context, if an element containing mixed directionality is not in the LTR context, that element will not be in the LTR context.

⚠ means you should check the test results, except for font differences.

Test results: Webfonts - Mozilla Firefox

### WOFF fonts

UA	IE	IE	Firefox	Chrome	Opera	Safari
version	8	7	3.6.3	4.1.249.1045	9.64	4.0.5
OS	XP, SP3					
OS renderer	Uniscribe 1.626.5756.0					
date	20100427	20100427	20100427	20100427	20100427	20100427
Alphabetic script (Georgian)	no ✓	no ✓	yes	no ✓	no ✓	no ✓
Alphabetic script (Armenian)	no ✓	no ✓	yes	no ✓	no ✓	no ✓
Complex scripts (Khmer)	no ✓	no ✓	yes	no ✓	no ✓	no ✓
Complex scripts (Hindi)	no ✓	no ✓	yes	no ✓	no ✓	no ✓
Shaping scripts (Arabic)	no ✓	no ✓	yes	no ✓	no ✓	no ✓
Shaping scripts (Urdu)	no ✓	no ✓	yes	no ✓	no ✓	no ✓
Mixed scripts (Thai)	no ✓	no ✓	yes	no ✓	no ✓	no ✓
Mixed scripts (Tibetan)	no ✓	no ✓	yes	no ✓	no ✓	no ✓
Mixed scripts (Myanmar)	no ✓	no ✓	yes	no ✓	no ✓	no ✓

Assertion:

Link to tests: WOFF fonts

### EOT fonts

UA	IE	IE	Firefox	Chrome	Opera	Safari
version	8	7	3.6.3	4.1.249.1045	9.64	4.0.5
OS	XP, SP3					
OS renderer	Uniscribe 1.626.5756.0					
date	20100427	20100427	20100427	20100427	20100427	20100427
Alphabetic script (Georgian)	no ✓					
Alphabetic script (Armenian)	no ✓					
Complex scripts	no ✓					

# Guidelines & education

## I18n resources

Articles

Tutorials

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

Tests

Talks

Tools

Reviews

The screenshot shows the W3C Internationalization (I18n) Activity website. The browser title is "W3C Internationalization (I18n) Activity - Mozilla Firefox". The address bar shows "http://www.w3.org/International/". The page content includes:

- Header:** W3C Internationalization (I18n) Activity, Making the World Wide Web world wide!
- Navigation:** Home Resources Techniques Topics News Groups About
- Home page:** The W3C Internationalization (I18n) Activity works with W3C working groups and liaises with other organizations to make it possible to use Web technologies with different languages, scripts, and cultures. From this page you can find articles and other resources about Web internationalization, and information about the groups that make up the Activity.
- Recent highlights:** A list of recent updates, including "Text direction tests updated" (30 April 10), "First translation into Hindi" (27 April 10), and "Webfonts tests updated to include WOFF" (27 April 10).
- Site links:** A sidebar menu with links to Getting Started, Tasks, Topics, Resource types, Reviews, Mail archives, Aggregated data, Translation, News by category, and News archives.
- I18n resources:** A sidebar menu with links to Authoring HTML & CSS, New! Authoring SVG, Authoring XML, Developing specifications, Setting up a server, Developing schemas, and Using the Web.
- Quick links:** A sidebar menu with links to Planet, Specifications, Articles etc., Tests, and @webi18n tweets.
- About the Activity:** A sidebar menu with links to Groups (Core, ITS, IG, JLTf), Mission, Contacts, and Activity Statement.
- Participate!** A sidebar menu with links to Join a Working Group, Review a W3C specification, Translate a specification or page, and Subscribe to the Interest Group list.
- Search for news:** A search box with radio buttons for "All Words", "Some Word", and "Entire phrase", and a "Search" button.
- Admin:** A sidebar menu with links to Admin, Profile (rishida), and Logout (rishida).

The main content area features an article titled "Text direction tests updated" dated 30 April 2010. The article text states: "The HTML and CSS text direction tests in the Internationalization Activity test suite have been updated as follows:"

- HTML5 and XHTML5 were added as new test formats
- tests now use strict DOCTYPEs
- reference graphics in the vertical text tests were improved
- the order of tests related to browser chrome was changed

In addition, a new results page was created for the tests about vertical text. Other changes behind the scenes include gathering of all tests into a single PHP file (generatehtml.php is no longer used). Code in the PHP file was also streamlined.

[search key: results-vertical-text]

Categories: Highlight, Text, w3cWebDesign, w3cWebUserAgents

The next article is "New translation into Hindi" dated 28 April 2010. The text says: "Thanks to the Bhavatmaj Seth, the following getting-started article has been translated into Hindi." The article title is "वेब पर भाषा (Language on the Web)".

[search keys: gs-language]

Categories: Translation, w3cWebDesign

The final article is "New translations into Spanish". The text says: "Thanks to the Spanish Translation Team, Spanish Translation US, the following articles have been translated into Spanish." The article titles are "Etiquetas que no están en inglés (Non-English tags)", "Ruby (Ruby)", and "Estilo y etiquetado de ruby (Ruby Markup and Styling)".

# Guidelines & education

# I18n resources

Articles

Tutorials

Technical notes

Tests

Talks

Tools

Reviews

## 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
- xml:lang in XHTML document schemas
- Declaring language in XHTML and HTML
- Setting language preferences in a browser
- Why use the language attribute?
- 2-letter or 3-letter language codes
- Choosing a language tag **NEW!**

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

### On this page

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

W3C I18n Site Index - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

http://www.w3.org/International/technique-index

W3C<sup>®</sup> Internationalization (I18n) Activity  
*Making the World Wide Web world wide!*

Home Resources Techniques Topics News Groups About

I18n site search:

RSS Feeds

## 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](#)

---

Author: Richard Ishida, W3C.

Content last changed 2009-03-12 16:23 GMT

Copyright © 2009 W3C<sup>®</sup> (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.

### 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 (eg. 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.

http://www.w3.org/International/technique-index?topic=htmlauth

# I18n resources

W3C I18n Site Index - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

http://www.w3.org/International/technique-index?topic=htmlauth

Google

**W3C** Internationalization (I18n) Activity  
*Making the World Wide Web world wide!*

Home Resources Techniques Topics News Groups About

I18n site search:

RSS Feeds

## 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
Select a general task...	Select a topic area...
<ul style="list-style-type: none"><li>▶ Using the Web</li><li>▶ Creating HTML &amp; CSS</li><li>▶ Authoring SVG</li><li>▶ Authoring XML</li><li>▶ Setting up a server</li><li>▶ Developing specifications</li><li>▶ Developing schemas</li></ul>	<ul style="list-style-type: none"><li>▶ Characters</li><li>▶ Language</li><li>▶ Markup &amp; text</li><li>▶ Text direction</li><li>▶ Styling &amp; layout</li><li>▶ Forms</li><li>▶ Navigation</li><li>▶ Cultural issues</li><li>▶ Troubleshooting</li></ul>

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.

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 (eg. 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.

http://www.w3.org/International/technique-index?topic=htmlauth

W3C Internationalization (I18n) Activity  
*Making the World Wide Web world wide!*

Home Resources Techniques Topics News Groups About

## 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...	Select a topic area...	Select a link...
<ul style="list-style-type: none"><li>▶ <a href="#">Using the Web</a></li><li>▶ <a href="#">Creating HTML &amp; CSS</a></li><li>▶ <a href="#">Authoring SVG</a></li><li>▶ <a href="#">Authoring XML</a></li><li>▶ <a href="#">Setting up a server</a></li><li>▶ <a href="#">Developing specifications</a></li><li>▶ <a href="#">Developing schemas</a></li></ul>	<ul style="list-style-type: none"><li>▶ <a href="#">Characters</a></li><li>▶ <a href="#">Language</a></li><li>▶ <a href="#">Markup &amp; text</a></li><li>▶ <a href="#">Text direction</a></li><li>▶ <a href="#">Styling &amp; layout</a></li><li>▶ <a href="#">Forms</a></li><li>▶ <a href="#">Navigation</a></li><li>▶ <a href="#">Cultural issues</a></li><li>▶ <a href="#">Troubleshooting</a></li></ul>	<ul style="list-style-type: none"><li><a href="#">Getting started</a></li><li><a href="#">Using attributes to declare language</a></li><li><a href="#">Declaring metadata about the language of the intended audience</a></li><li><a href="#">Declaring language for documents aimed at speakers of more than one language</a></li><li><a href="#">Choosing language values</a></li><li><a href="#">Identifying in-document language changes</a></li><li><a href="#">Indicating the language of a link destination</a></li><li><a href="#">Styling by language</a></li><li><a href="#">Using Accept-Language for locale setting</a></li></ul>

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.

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 (eg. 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.

The screenshot shows a Mozilla Firefox browser window with the address bar displaying <http://www.w3.org/International/techniques/authoring-html#langvalues>. The page content is as follows:

## 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 zxx 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 4647 (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.

Choose another topic ▶  
Go to the techniques index ▶

---

## Identifying in-document

### Recommendations

- Use the `lang` and/or `xml:lang` attributes around text to indicate any

The screenshot shows a Mozilla Firefox browser window with the address bar displaying <http://www.w3.org/International/questions/qa-choosing-language-tags>. The page title is "Choosing a Language Tag" under the W3C Internationalization header. The page content includes a navigation menu (Home, Resources, Techniques, Topics, News, Groups, About), a search bar, and a "related links" sidebar with items like Resources, Techniques, IANA Language Subtag Registry, Subtag search tool, and BCP 47. The main text is organized into sections: "question", "background", and "answer".

**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 8,000 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.

BCP stands for 'Best Current Practice', and is a persistent name for a series of RFCs whose numbers change as they are updated. The latest RFC describing language tag syntax is RFC 5646, Tags for the Identification of Languages, and it obsoletes the older RFCs 4646 3066 and 1766.

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

Note, also, that some environments or systems may dictate choices that are different from what you would otherwise expect. For example, in Java you must use "hw" (deprecated in BCP47) in place of "he" (recommended in BCP47).

# Text expansion

**DochuLa Pass**



Around 3,100 m high, in Bhutan.

©  Anyone can see this photo ([edit](#))

Uploaded on Jan 28, 2007 | [Map](#) | [Delete](#)

1,414 views / 5 comments

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

# Text expansion

## Seals and sealions at Pier 39

ADD NOTE SEND TO GROUP ADD TO SET BLOG THIS ALL SIZES ORDER PRINTS ROTATE DELETE X



## Seals and sealions at Pier 39

ADICIONAR OBSERVAÇÃO ENVIAR PARA UM GRUPO ADICIONAR AO ÁLBUM BLOGAR TODOS TAMAÑHOS IMPRIMIR GIRAR EXCLU



# Text expansion

## Global settings

Interface language

Search language

Number of results

## Acuan Umum

Bahasa Pengantar di Antar Muka

Bahasa Pengantar untuk Penelusuran

Jumlah Hasil Penelusuran

## Allgemeine Voreinstellungen

Sprache der Benutzeroberfläche

Suchsprache

Anzahl der Ergebnisse

# Guidelines & education

## Checker tool

▼ Address http://rishida.net/tools/i18nchecker/test.php

http://rishida.net/tools/i18nchecker/test.php

Run another check

▼ Results

 **2**
 **8**
 **1**

▼ Information XHTML 1.0 :: text/html

Character encoding		Code
HTTP Content-Type	No charset found.	Content-Type: text/html
Byte order mark (BOM)	UTF8	
xml declaration	None found.	
meta charset element	iso-8859-1	<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
HTML5 meta charset element	None found.	
Language		Code
<html lang=	kk	<html lang="kk" xml:lang="to" dir="ltr" xmlns="http://www.w3.org/1999/xhtml">
<html xml:lang=	to	<html lang="kk" xml:lang="to" dir="ltr" xmlns="http://www.w3.org/1999/xhtml">
HTTP Content-Language	ka, ta	Content-Language: ka, ta
meta content-language element	en,fr,sp	<meta http-equiv="Content-Language" content="en,fr,sp" />
Text direction		Code
Default direction	ltr	<html lang="kk" xml:lang="to" dir="ltr" xmlns="http://www.w3.org/1999/xhtml">
Class & id names		Code
Non-ascii class or id names	8	Show list
Non-NFC class or id names	4	Show list
Request headers		
Accept-Language	en,pt-br;q=0.8,fr;q=0.5,ch;q=0.3	
Accept-Charset	ISO-8859-1,utf-8;q=0.7,*;q=0.7	

# Guidelines & education Checker tool

Accept-Charset ISO-8859-1,utf-8;q=0.7,\*;q=0.7

↑ TOP

## ▼ Detailed report

Severity	Description
----------	-------------



### Multiple encoding declarations using the meta tag.

- `<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />`
- `<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.

↑ TOP



### Incorrect values used for dir attribute.

- `<p dir="rlt" lang="he" xml:lang="he" title="Hebrew : Hebrew" class="phrase">`

The value should be one of 'rtl' or 'ltr'.

↑ TOP



### 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, ie. be normalised. We recommend Unicode Normalized Form C (NFC).

Total number of non-NFC names: 4.

- `<p title="Latin : Hungarian" lang="hu" xml:lang="hu" class="phrase Nemzetköziesítés">`
- `<p title="Latin : Welsh" lang="cy" xml:lang="cy" class="phrase" id="Teruïet">`
- `<div id="Ā&π">`
- `<p id="cãn" >`

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

↑ TOP



### ▶ 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](mailto:www-international@w3.org)), and track other technologies for internationally relevant topics. Follow our RSS feeds and twitter channels ([@webi18n](https://twitter.com/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/>