



Practical & Cultural Issues in Web Design

Richard Ishida
W3C Internationalization Activity Lead

Objectives

After this tutorial you should be able to say:

- ▶ identify a number of practices that affect easy localization of your Web pages
- ▶ why you need to think about localization earlier than people typically expect
- ▶ that you need to learn more about internationalization

Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 2



Terminology

Localization

The **adaptation** of a product, application or document content to meet the language, cultural and other requirements of a specific target market.

Internationalization

The **design and development** of a product, application or document content that **enables** easy localization for target audiences that vary in culture, region, or language.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 3

Localization without internationalization can be very hard. This presentation will use examples to make that point, and stress the value of considering internationalization as an integral part of the design and development activity – not an afterthought left to the 'localization folks'.



Separating content & presentation

Separating content & presentation

- Linguistic issues
- Cultural issues
- Internationalization recap
- Getting help

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 4

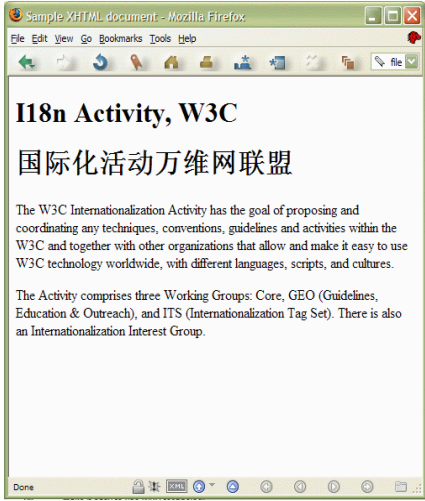
W3C®

Content (XHTML)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">


<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
<title>About the W3C I18n Activity</title>
<style type="text/css" src="mystyling.css" />
</head>

<body>
<h1>I18n Activity, W3C</h1>
<div class="international-text" xml:lang="zh-Hans"
      lang="zh-Hans">国际化活动万维网联盟</div>
<div class="description">
<p>The W3C Internationalization Activity has the goal of proposing
and coordinating any techniques, conventions, guidelines and
activities within the W3C and together with other organizations
that allow and make it easy to use W3C technology worldwide,
with different languages, scripts, and cultures.</p>
<p>The Activity comprises three Working Groups: Core, GEO
(Guidelines, Education & Outreach), and ITS (Internationalization
Tag Set). There is also an Internationalization Interest Group.</p>
</div>
</body>
</html>
```



Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 5

This slide shows some bare HTML content, displayed using the browser's default styling.



Separating content & presentation
 The basics

Content (XHTML)

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
<title>About the W3C I18n Activity</title>
<style type="text/css" src="mystyling.css" />
</head>

<body>
<h1>I18n Activity, W3C</h1>
<div class="international-text" xml:lang="zh-Hans"
      lang="zh-Hans">国际化活动万维网联盟</h1>
<div class="description">
  <p>The W3C Internationalization Activity has the goal of proposing
    and coordinating any techniques, conventions, guidelines and
    activities within the W3C and together with other organizations
    that allow and make it easy to use W3C technology worldwide,
    with different languages, scripts, and cultures.</p>
  <p>The Activity comprises three Working Groups: Core, GEO
    (Guidelines, Education & Outreach), and ITS (Internationalization
    Tag Set). There is also an Internationalization Interest Group.</p>
</div>
</body>
</html>

```

Presentation (CSS)

```

body {
  background: white;
  color: black;
  font-family: serif;
  font-size: 1em;
}

h1 {
  font-size: 240%;
}

div.international-text {
  font-family: MingLiu, sans-serif;
  font-size: 240%;
}

p {
  margin-top: 1em;
}

```

Copyright © 2005 W3C (MIT, ERCIM, Keio)
 slide 6

The HTML is shown on the left. There is no presentational information in the HTML – which is as it should be.

To the right is some CSS code that applies styling to the HTML.

W3C®

Three overlapping Mozilla Firefox browser windows are shown, each displaying the same HTML content but with different CSS styling. The windows illustrate how the same content can be presented differently by changing the CSS file.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 7

Each of these windows shows EXACTLY the same HTML file. The changes made to the CSS file produced three very different presentations of that basic content.

This is particularly useful for changing the presentational aspects of a site or group of pages. You typically only need to edit a single CSS file, rather than editing all the code of each HTML file.

This can also be beneficial for localization, since typographic approaches, colors, etc, may need to be changed for different locales. Making such changes in the CSS is much easier than adapting the HTML.

Separating content & presentation

The basics

W3C®



Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 8

Remember, also, that the Mobile Web is becoming increasingly important these days – and may be especially so in developing countries in the future. This means that content needs to be adapted to fit on handheld devices with smaller screens.

Again, this would ideally be achieved by styling the content, rather than writing a completely separate Web.

You should not make assumptions, when creating content, that you know what it will look like when finally displayed. These days, it may well be displayed in a number of different formats.

Separating content & presentation

International issues

W3C®


- ◆ problems of resolution to support bold and italics in small CJK characters on-screen
- ◆ different ways of emphasizing text in Japanese (wakiten & amikake)

これは日本語です。

これは日本語です。


Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 9

Here are some ways in which typographic differences may appear between language versions of the same content.



Separating content & presentation

International issues



- ◆ problems of resolution to support bold and italics in small CJK characters on-screen
- ◆ different ways of emphasizing text in Japanese (wakiten & amikake)
- ◆ no upper- vs. lower-case distinction in most non-Latin scripts
- ◆ no convention of distinguishing between proportional and mono-spaced fonts for some scripts

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 10

Separating content & presentation
Practical implications


W3C®


Making the World Wide Web *worldwide*.

<p>Making the World Wide Web <i>worldwide</i></p>

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 11

You should try to remove all presentational constructs from your content. For example, use of `<i>` tags shows that you are assuming that the text will be italicized. Because ideographic text doesn't support italicizations well in small font sizes, you could be causing problems for localization.

 Separating content & presentation
Practical implications



Making the World Wide Web *worldwide*.


<p>Making the World Wide Web worldwide</p>



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 12

Not only is it better for localization to express the idea or semantics in the content, and leave the presentation to the style sheet, it will also improve your original text by making you more aware of what you are actually doing.



Separating content & presentation
Practical implications

See the **System Administrator Guide** for an example of re-use.

<p>See the System Administrator Guide
for an example of re-use.</p>

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 13

The same applies to document conventions such as representation of referenced resources. When using class annotations or microformats, don't describe the expected presentational rendering, describe the function of the text.

Separating content & presentation

Practical implications

W3C®

See the **System Administrator Guide** for an example of re-use.

<p>See the System Administrator Guide
for an example of re-use.</p>

doctype
chaptertitle
inputsequence
etc.

slide 14

Copyright © 2005 W3C (MIT, ERCIM, Keio)

Setting document direction

How to mark up the document

W3C®

[Home](#) | [Tags](#) | [Groups](#) | [People](#) | [Invite](#)

Logged in as [r12a](#) | [Your Account](#) | [Help](#) | [Sign Out](#)

Photos: [Yours](#) • [Upload](#) • [Organize](#) • [Your Contacts](#) • [Explore](#)

[flickr](#) BETA

Château de La Napoule

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

Uploaded on June 9, 2005
by [r12a](#)



[r12a's photostream](#)
1177 photos
[View as slideshow](#)

Tags
 0506-cote-dazur [x]
 Mandelieu-La Napoule [x]
 Alpes-Maritimes [x]
 France [x]
[Add a tag](#)

Additional Information
 © All rights reserved ([change](#))
 Taken with a Fujifilm FinePix S7000.
[More properties](#)
 Taken on [June 6, 2005](#) ([edit](#))
 See [different sizes](#)
 Viewed 92 times. (Not including you)
[Edit](#) title, description, and tags
[NEW](#) [Replace](#) this photo

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 15



Linguistic issues

- Separating content & presentation
- Linguistic issues
- Cultural issues
- Internationalization recap
- Getting help

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 16

Linguistic issues
Syntax differences

W3C®

They are speaking to her from my new house.


Están hablándole desde mi casa nueva.

私の新しい家から彼女と話しています。

تكلّمونها من بيتي الجديد

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 17

This slide shows the same idea expressed in multiple languages. Within each translation of the sentence, the number of words is different, and the order of those words changes.



Linguistic issues
Agreement



Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 18

A word in one language may map onto more than one word in another, as seen in the Spanish translations for 'on'. In Spanish it depends what is on.

In addition, Spanish uses different endings according to whether the thing that is on is masculine, feminine, singular or plural.

Linguistic issues
Plurality


W3C

Number of pages	Russian word for 'page'
1	страниц ^а
2-4	страниц ^ы
5-10	страниц
11-20 (irregular)	страниц
21	страниц ^а
22-24	страниц ^ы
25-30	страниц
> 30	Repeat pattern of endings for 1 to 10.


Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 19

Plurality is much more complicated in Arabic and Russian than it is in English. The shape of the Russian word for 'page' changes according to the number in a (mostly) cyclical way.



Linguistic issues
Text expansion



Input processing
features

Eingabeverarbeitungsfunktionen

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 20

Text in English and Chinese is typically much more concise than text in other languages. In some languages, such as Finnish and German, this tendency is worse in short English noun phrases that become a single word in translation.

Linguistic issues

Sentence-like composite messages

W3C®

The < > has been disabled.

printer

stacker

stapler options

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 21

The diagram illustrates a linguistic issue in a user interface. At the top, a blue header bar contains the text 'Linguistic issues' and 'Sentence-like composite messages' on the left, and the 'W3C®' logo on the right. Below this, the sentence 'The < > has been disabled.' is displayed in blue. Three yellow arrows point from the words 'printer', 'stacker', and 'stapler options' (also in blue) to the angle brackets '<' and '>' in the sentence. 'printer' points to '<', 'stacker' points to '>', and 'stapler options' points to the space between them. This visualizes a developer's attempt to reuse a common sentence structure to save memory, which results in an untranslatable phrase due to linguistic rules about gender and number agreement in many languages.

In this example, the developer has tried to save memory by re-using part of a common sentence. Unfortunately, because of the effects of rules about agreement between gender and number in many languages, this becomes an untranslatable phrase. The developer needs to be aware of the likely impact on translatability of such things.

Linguistic issues

Topic-comment composite messages

W3C®

Printer: ☒ enabled

Stacker: ☒ enabled

Stapler options: ☐ enabled

Imprimante: ☒ activée

Module de reception: ☒ activé

Options d'agrafage: ☐ activées

<http://www.w3.org/International/articles/composite-messages/>

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 22

Composite messages in topic-comment arrangements typically cause no issues for translation (as long as you don't re-use the comment string in multiple contexts).

Linguistic issues
Text string re-use

W3C®

```
<?php
function getStatus (device) {
    if ($device.status != 0) { return 'On'; }
    else { return 'Off'; }
}

?>
...
<p>Printer: <?php print getStatus($printer); ?></p>
<p>Stacker: <?php print getStatus($stacker); ?></p>
<p>Stapler options: <?php print getStatus($soptions); ?></p>
```

<http://www.w3.org/International/articles/text-reuse/>

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 23

This shows an example where text-string re-use will cause significant localization problems, due to the word mapping and agreement features of many languages.

The code prevents the use of multiple translations for the words 'on' and 'off'.


Linguistic issues


Variable variable positions

W3C®

There were %d spelling mistakes in file %s.

Datei %s enthält %d Rechtschreibfehler.

`printf("There were %d spelling mistakes in file %s.",
currentpage, totalpages)` 

`printf("There were %1$d spelling mistakes in file %2$s .",
currentpage, totalpages)` 

`printf("Datei %2$s enthält %1$d Rechtschreibfehler.",
currentpage, totalpages)`

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 24

This is an example of syntax differences affecting development techniques.

The order of variables needs to be different between English and German versions. Unless you are using slightly more advance techniques in PHP, you will prevent this possibility and seriously affect translatability.

W3C®
Linguistic issues
Managing expansion space

Interface Language

Sprache der Benutzeroberfläche

Interface Language

Sprache der Benutzeroberfläche

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 25

English and Chinese text usually expand when translated. You should consider the potential impact of this on page design, and either allow text to flow into larger areas, or leave expansion space.

For example, putting labels beside form fields is often likely to cause expansion space problems. This issue can often be avoided by allowing text to expand above the field, instead.





Cultural issues

Separating content & presentation	Formatting & reading data
Linguistic issues	Symbolism & examples
Cultural issues	Color
Internationalization recap	Differing approaches
Getting help	International text

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 26

Setting document direction

Formatting & reading data

Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 27

Cultural issues

Formatting & reading data

W3C®

03/02/01
2.3.01
2001-III-2
2001年3月2日

Date:*

Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 28

The expected format of data such as dates and times varies around the world. You should internationalize your deliverable so that it respects local customs.

Much more difficult than displaying a date correctly, however, is recognizing a date supplied by a user if you don't clearly signal the expected format.

W3C[®]

Cultural issues
Formatting & reading data

Chinese	Chang Wen Yun
Malay	Isa (bin) Aman
Indian	M. Thiruselvan
	Thiruselvan s/o Manickavasagam

Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 29

People's name formats can also be very different around the world. It is important to consider how you will ask for people's names in form design, but also how you will capture the information in your database.

Cultural issues

Formatting & reading data

W3C®

Россия
г. Пермь 614055
ул. Крупской 93-82
Селивановой Юлии

Country:

United States

First name:

Last name:

Address:

City:

State:

AZ

Zip code:

Telephone:


()

Application date:

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 30

Be careful about assuming what others' name and address formats will be. Also think about how you will store the names and addresses in the database. For example, do you really need to split out street number? How will you generate a Russian or Japanese address that goes from general to specific from top to bottom?


W3C[®]

Cultural issues


Formatting & reading data

- Numeric formats
- Calendars & date formats
- Address formats
- Name conventions
- Time formats
- Currency formats
- List separators
- Telephone numbers
- etc....

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 31



Setting document direction



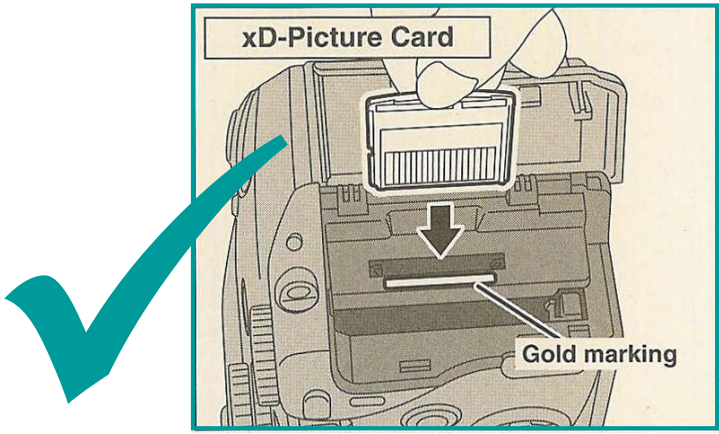
Symbolism & examples

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 32

Cultural issues
Symbolism & examples

W3C®



The diagram shows a top-down view of a camera's memory card slot. An xD-Picture Card is being inserted into the slot. A large teal checkmark is superimposed on the left side of the diagram. A label 'xD-Picture Card' points to the card. Another label 'Gold marking' points to a small gold-colored area on the card. The text 'slide 33' is in the bottom right corner of the slide.

Copyright © 2005 W3C (MIT, ERCIM, Keio)

Symbolism can differ from place to place. For example the check mark means *incorrect* in some places around the world.

Ensure that you do not give the wrong message through your use of colors, symbolism, examples, etc.



Symbolism can differ from place to place. For example the check mark means *incorrect* in some places around the world.

Cultural issues
Symbolism & examples

W3C®




Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 35

Graphics may need to be changed if they don't reflect the local culture of certain places.

Cultural issues

Symbolism & examples


W3C®




Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 36

Body language and gestures are particularly dangerous. Each of these symbols can give offense in one part of the world or another.



Setting document direction



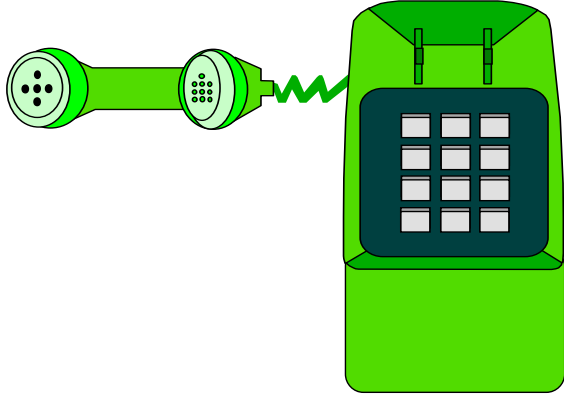
Color

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 37

Cultural issues
Color

W3C®



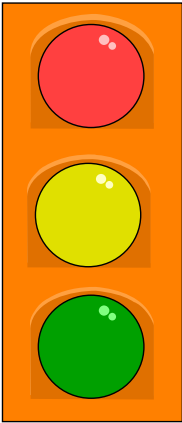
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 38

Color can have strong subliminal suggestiveness due to conventional uses in particular cultures.

Cultural issues

Color

W3C®




Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 39

Also, color names can change from culture to culture. Germans, for example, don't have a strong concept of the color 'purple'.

Cultural issues
Color

W3C®




Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 40

Color also has different connotations in different parts of the world.



Cultural issues
Color

W3C®



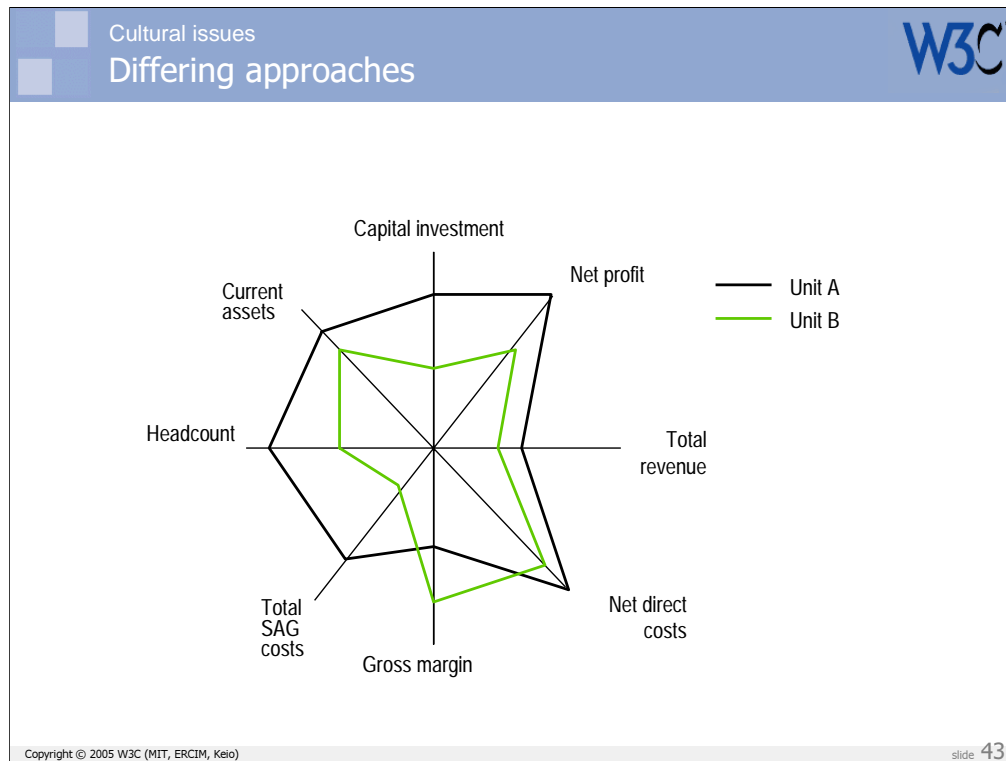
Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 41

Setting document direction

Differing approaches

Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 42



Then you need to be aware that people in different parts of the world may do things in different ways. For example, the radar chart was such a common way of representing comparative data in Japan that, when Lotus 1-2-3 was launched in that area they had to reengineer it to add that.

Cultural issues
Differing approaches

W3C®

"... one Latin American teacher recently complained to me that the US-manufactured and well-translated educational software currently being used in his country's primary schools presupposed 'solitary problem solvers', whereas his culture stressed collective problem-solving."

Kenneth Keniston,
Language International, May 1996

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 44

Considerations of this kind require you to make big decisions at the very start of the development phase about how to proceed. Otherwise you could waste a lot of time and energy producing something that doesn't meet your customer's needs.

Cultural issues

Differing approaches

W3C

YAHOO! UK & IRELAND

Yahoo! Travel Plan your May Bank Holiday getaway! Cruise the Med What is Fórum Barcelona 2004?

Website Directory Sites organised by subject.

Arts & Humanities
[Literature](#), [History](#), [Photography](#)...

Business & Economy
[B2B](#), [Shopping](#), [Investments](#),
[Property](#)...

Computers & Internet
[Internet](#), [Reviews](#), [Software](#),
[Games](#)...

Education
[UK](#), [Ireland](#), [Universities](#)...

Entertainment
[Humour](#), [Movies](#), [Music](#), [Actors](#)...

Government
[UK](#), [Ireland](#), [Politics](#), [Law](#)...

Health
[Medicine](#), [Drugs](#), [Diseases](#), [Fitness](#)...

News & Media
[Newspapers](#), [Weather](#), [TV](#)...

Recreation & Sport
[Sport](#), [Hobbies](#), [Travel](#), [Motoring](#)...

Reference
[Maps](#), [Dictionaries](#), [Phone Numbers](#)...

Regional
[UK](#), [Ireland](#), [Countries](#)...

Science
[Animals](#), [Geography](#), [Engineering](#)...

Social Science
[Economics](#), [Languages](#), [Psychology](#)...

Society & Culture
[People](#), [Food & Drink](#), [Environment](#),
[Sexuality](#)...

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 45

This and the following slides show how Yahoo adapts its categorizations to reflect the preoccupations of various different countries.

Cultural issues
W3C®

Differing approaches

Guide Web - Classement thématique de sites Web

[Suggérer un site](#)

<p><u>Actualités et médias</u> Journaux, Télévision, Météo...</p> <p><u>Commerce et économie</u> B2B, Shopping, Emploi, Immobilier...</p> <p><u>Informatique et Internet</u> Internet, Logiciels, Fonds d'écran...</p> <p><u>Santé</u> Diététique, Médecine, Thalasso...</p> <p><u>Enseignement et formation</u> Primaire, Secondaire, Supérieur...</p> <p><u>Institutions et politique</u> Ministères, Droit, Politique...</p> <p><u>Sciences et technologies</u> Animaux, Astronomie, Physique...</p>	<p><u>Sports et loisirs</u> Foot, Tourisme, Auto/Moto, Jeux...</p> <p><u>Art et culture</u> Littérature, Cinéma, Musique, BD...</p> <p><u>Divertissement</u> Tests/Quiz, Loteries, Humour, Sorties...</p> <p><u>Classement géographique</u> Pays, Europe, France, Paris...</p> <p><u>Références et annuaires</u> Dictionnaires, Annuaire, Cartes/Atlas...</p> <p><u>Société</u> Enfants, Gastronomie, Rencontres...</p> <p><u>Sciences humaines</u> Archéologie, Histoire, Psychologie...</p>
--	--

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 46



Cultural issues
Differing approaches

W3C®

Yahoo! カテゴリ サイトの推薦


エンターテインメント 映画, 音楽, 芸能人, コミック, 占い ...	メディアとニュース テレビ, ラジオ, 新聞, 雑誌 ...
趣味とスポーツ アウトドア, ゲーム, 車, スポーツ, 旅 ...	ビジネスと経済 ショッピング, B2B, 雇用, 金融 ...
芸術と人文 写真, 建築, 美術館, 歴史, 文学 ...	各種資料と情報源 図書館, 辞書, 郵便, 電話番号 ...
生活と文化 子ども, 環境, グルメ, 障害者 ...	コンピュータとインターネット ハードウェア, ソフトウェア, WWW ...
教育 大学, 専門学校, 小中高, 資格 ...	政治 政治, 行政, 国会, 法 ...
健康と医学 病院, 病気, ダイエット ...	自然科学と技術 動物, エコロジー, 地球, 天文, 工学 ...
社会科学 経済学, 社会学, 言語, 政治学 ...	地域情報 日本の地方, 世界の国 ...

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 47

Setting document direction

International text

Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 48


W3C®

Cultural issues
International text

The outcome was a heavy defeat for the new House Democratic leadership. Ways and means committee chairman Dan Rostenkowski looked particularly inept. He took a vacillating, off-again, on-again, stand and eventually lost control of his committee. Majority Leader Richard Gephardt, failed to recognize the strength of the drive for a capital-gains cut. Finally, but too late, he helped draft the Democratic alternative - which combined deductible IRA contributions for everybody with an increase in the tax rate on people with incomes over \$200,000. The capital gains cut, he declared, was "designed to keep Leona Helmsley's dream alive - that only little people pay taxes." Republicans retorted, in effect, There you go again, proposing a tax increase. "Fish gotta swim, birds gotta fly, and Democrats gotta raise taxes" was the way one G.O.P. quip put it.

Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 49

The text in blue is difficult for a British or non-English speaking person to understand. Sometimes because of references to things that don't exist in other cultures, sometimes because of culture-specific knowledge.

W3C®

Cultural issues
International text

Montre qui tu es !

The 5 R's

Nothing sucks like an Electrolux!

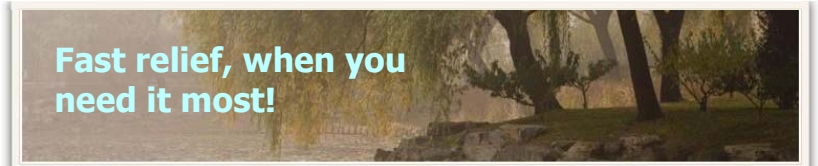
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 50

You should check your text for things that might be interpreted differently in another culture.

Cultural issues

International text

W3C®



**Fast relief, when you
need it most!**

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 51

When dealing with graphics, consider how to deal with text. Ideally the text will be overlaid on a graphic, rather than embedded in it. If the text is within the graphic, try to ensure that you develop it in layers, with text on a separate layer, so that when it comes to translation the text can be easily removed and replaced over complicated backgrounds.



Internationalization recap

- Separating content & presentation
- Linguistic issues
- Cultural issues
- [Internationalization recap](#)
- Getting help

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 52





Internationalization recap

Internationalization means:

- using a Quality approach to reduce the overall cost and time to market/release of multinational deliverables
- **designing** into the deliverable or product an internationalized base, and a modular and easily adaptable architecture
- not always doing extra work – maybe just working in a better way

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 53




Getting help

- Separating content & presentation
- Linguistic issues
- Cultural issues
- Internationalization recap
- Getting help

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 54

Getting help
GEO (Guidelines, Education & Outreach) WG

How do I ...

- Ensure that XHTML forms return data in the right encoding?
- Make my Urdu, Arabic or Hebrew text display correctly?
- Declare language and encoding for XML documents?
- Order XSL output according to French rules?
- Approach the creation of multilingual documents in HTML?
- Help users navigate to the right localized page?
- Ensure the table I'm about to write has all the right i18n features?
- etc

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 55

Getting help

I18n Activity home page

W3C®

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

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 56

A range of helpful materials are available from the Internationalization home page.

Getting help

Topic & techniques indexes

W3C®

Language

- lang
- Content-Language
- browser settings
- declaring
- flags
- Hans and Hant language codes
- hreflang
- IANA language tags
- ISO language codes
 - ISO 639 Codes for the Representation of Names of Languages
 - 2-letter and 3-letter codes
 - Article: Two-letter or three-letter language codes
- ISO country codes
- ISO script codes
- language tag values
- language negotiation
- link target language
- primary language
- RFC 3066
- RFC 3066bis
- text processing language
- xml:lang

Back to contents

Markup & text

- composite messages & string re-use
- non-English markup
- ruby

Back to contents

Choosing language values

- How to choose language values
 - W3C techniques document (Building Techniques for XHTML & HTML, Internationalization: Specifying the language of content)
 - Language tags in XHTML and XML
 - How to choose the right attribute values
- Specifying language tag values
 - (W3C tutorial: Declaring Language in XHTML and HTML)
- Two-letter or three-letter language codes
 - Should I use two-letter or three-letter language codes? (W3C article)
- Best practices
 - Follow the guidelines in RFC3066 or its successors for language attribute values
 - Use the two-letter ISO 639 codes for the language code where there are both 2- and 3-letter codes
 - UK: none
 - Where possible, use the codes *en-GB* and *en-US* to refer to Simplified and Traditional Chinese, respectively
 - UK: none
- Particularly useful links
 - IANA Assigned Language Tags
 - IANA language tag registry
 - RFC 3066 Tags for the identification of Languages
 - The IETF document that defines how to use language tags to identify languages
 - ISO 3166: Codes for Country Names
 - ISO country codes
 - ISO 639: Codes for the Representation of Names of Languages
 - ISO language codes
- Other references
 - Specifying the language of content: the lang attribute
 - lang in the HTML 4.01 spec (section 5.1)
 - Language identification
 - and lang in the XML spec (section 2.12)
- Test data
 - Automatic font assignment for CJK test
 - W3C test page
 - Automatic font assignment for CJK test
 - W3C test results

Back to technique index

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 57

We are building indices to help you locate the information more easily.

Getting help

Articles, tutorials, tests, etc.

W3C®

language tags in HTML and XML - Author's Preview

W3C Architecture domain Internationalization

Home About Groups Topics Techniques Resources News

Language tags in HTML and XML

on this page: [rfc 3066](#) - [special primary subtags](#) - [iana-registered tags](#) - [matching tags](#) - [issues with tags](#) - [by the way](#) - [further reading](#)

Language tags can be (and should be) used to indicate the language of text in HTML and XML documents. For HTML, 4 language tags are specified with the lang attribute. For XML, language tags are given in the xmlns:lang attribute. In both cases, language information is inherited along the document hierarchy, i.e. it has to be given only once if the whole document is in one language, and language information nests, i.e. inner attributes overwrite outer attributes.

Language tags are defined in RFC 3066, which obsoletes the older RFC 1766. XML has been updated to use RFC 3066 by an erratum. RFC 3066 is based on ISO-639 two-letter and three-letter language codes, and on ISO-3166 two-letter country codes. RFC 1766 did not include three-letter language codes.

Examples include:

Code	Language	Explanation
en	English	ISO-639 two-letter language code
mis	Malay	ISO-639 three-letter language code
fr-CA	French as used in Canada	ISO-639 two-letter code with ISO-3166 two-letter country code
en-scouse	English Liverpoolian dialect known as 'Scouse'	ISO-639 two-letter language code with addition, IANA-registered
x-klingon	Klingon	IANA-registered language code
x-pig-latin	Pig Latin	Unregistered/Experimental

Language tags starting with x- are defined in the IANA registry of language tags. Language tags starting with x- denote experimental tags without guarantee for uniqueness. The list of ISO-639 two-letter and three-letter language codes is provided by the ISO 639-2 Registration Authority (Library of Congress, USA).

According to RFC 3066, for languages with both a two-letter and a three-letter code, the


related links


[RFC 3066: Tags for the identification of Languages](#)
[ISO 639: Codes for the Representation of Names of Languages](#)
[ISO 3166: Codes for Country Names](#)
[IANA language tag registry](#)
[Authoring Techniques for XHTML & HTML, Internationalization: Specifying the language of content 1.0](#)
[W3C HTML resource index: Language declarations and language negotiation](#)


Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 58

An example of an article.

 Getting help

 Getting involved




Get involved:


- visit the I18n Activity Home Page
- join a W3C Internationalization Working Group, or the Interest Group (www-international@w3.org)
- offer to help with reviews, or provide local knowledge for other WGs
- provide translations of W3C specifications or articles
- take advantage of the i18n-readiness of W3C technology

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 59



Getting help
And finally...



- ◆ this is your Web – not the W3C's – if something isn't right, get involved to fix it

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 60