




Designing for International Users: Practical Tips

Richard Ishida
W3C Internationalization Activity Lead

Online version available from
<http://www.w3.org/2007/Talks/0706-atmedia/>

W3C[®]

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 2



Linguistic issues

Linguistic issues

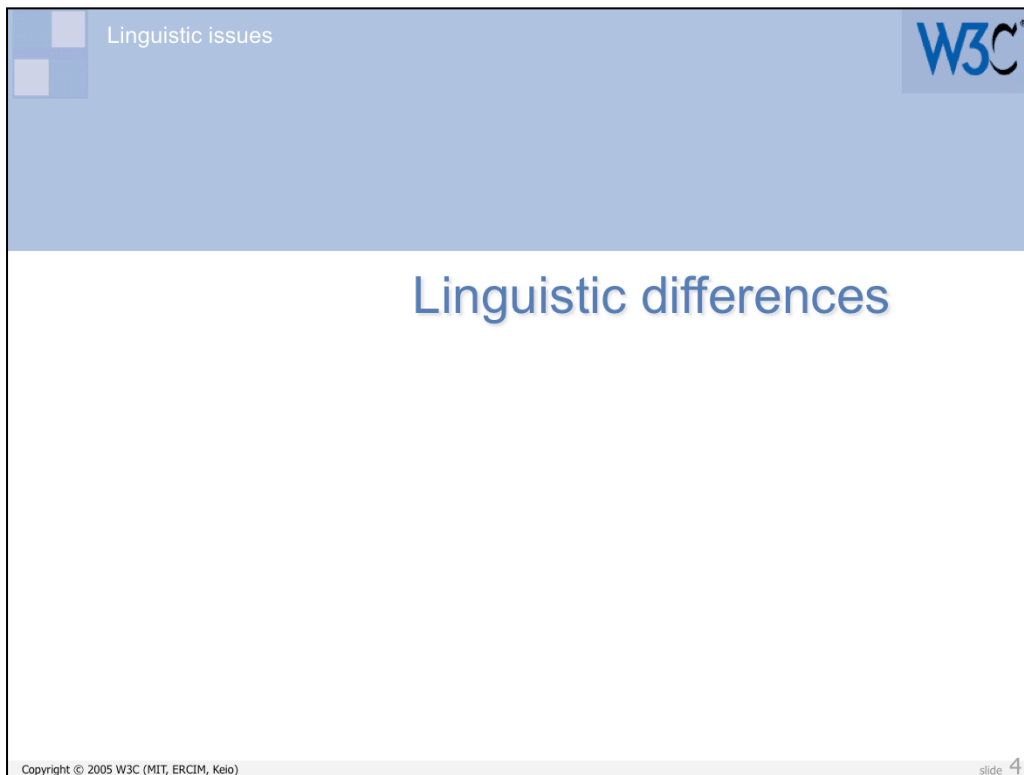
- Separating content & presentation
- Cultural issues
- What this means to me

Linguistic differences

- Composite messages
- Managing text expansion

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 3




Linguistic issues

W3C®


Linguistic differences

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

In this first section we will look at a few ways in which languages differ, and then see those differences causing practical issues for localization where the developer/designer has not thought about internationalization.



Linguistic issues
Syntax differences



You are speaking to her from my new house.

Estáis 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 5

This shows four different ways of writing one idea. In each case the order of 'words' and the number of 'words' is different.

Linguistic issues
Agreement & Concept mappings

W3C®

On

Conectado	-os -a -as
Encendido	-os -a -as
Activado	-os -a -as

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

This slide shows how the English word 'On' can map to three different words in Spanish. And then there are the masculine, feminine and plural forms of agreement that change the shape of the word according to its context.

Linguistic issues
Agreement

W3C

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

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 7

In Russian there is a complex plural system. Apart from the irregular teens, the word endings are applied in a rotating way.

Linguistic issues
Text expansion



Trim

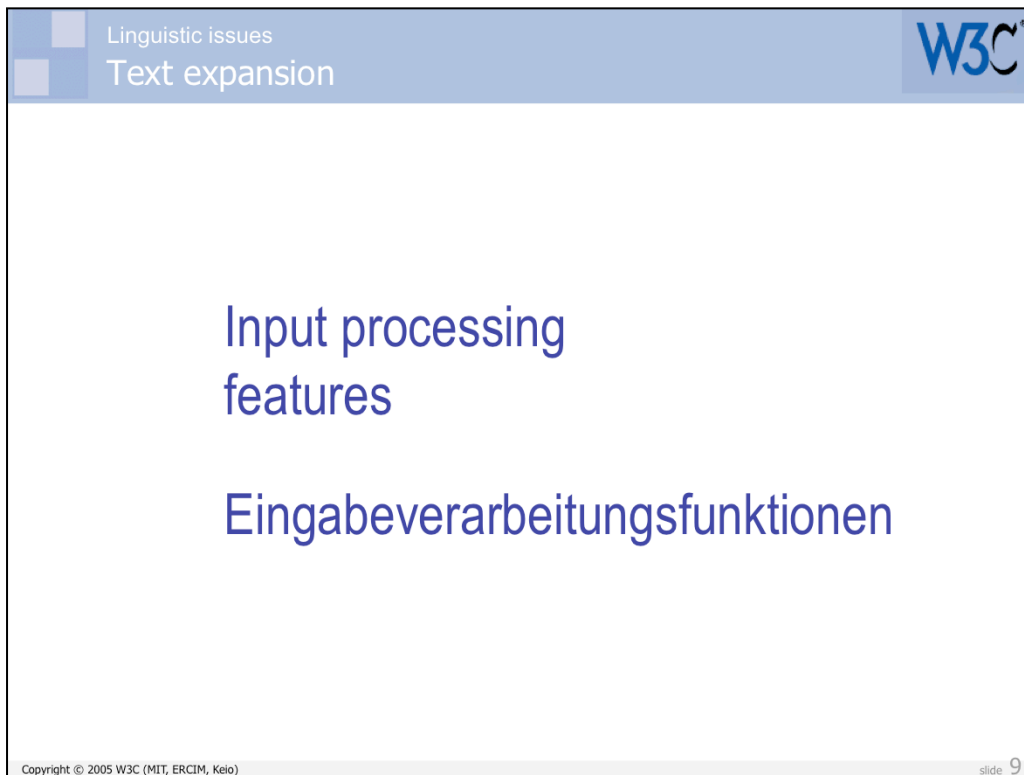
Kantbeskæring

Effacement de bords

Eliminación de bordes

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

This slide introduces the idea that terms or labels can be of widely differing lengths in different languages.



The slide features a blue header bar with the text 'Linguistic issues' and 'Text expansion' on the left, and the 'W3C' logo on the right. The main content area is white and contains the text 'Input processing features' and 'Eingabeverarbeitungsfunktionen' in blue. The footer is a light gray bar with copyright information on the left and 'slide 9' on the right.

Linguistic issues
Text expansion


W3C[®]

Input processing
features


Eingabeverarbeitungsfunktionen

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

In languages such as German, Dutch or Swedish it is common to find English 'compound nouns' expressed as a single, long word.



Linguistic issues



Composite messages

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 10

Linguistic issues
Handling variables
W3C®

W3C®
QUALITY Assurance
Markup Validation Service v0.7.0

Home
About...
News
Docs
Help & FAQ
Feedback

Result: Failed validation
There were 268 validation errors in the file myFirst.html

Jump To:
Results

File: webdeveloper_localhost_1159293639132.html
Modified: (undefined)
Server: Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.8.0.7) Gecko/20080909 Firefox/1.5.0.7
Size: (undefined)
Content-Type: text/html
Encoding: utf-8
Doctype: XHTML 1.0 Transitional
Root Namespace: http://www.w3.org/1999/xhtml
Note: The Validator XML support has some limitations.

This page is **not** Valid XHTML 1.0 Transitional!

Below are the results of checking this document for XML well-formedness and validity.

1. **Error** Line 14 column 103: end tag for "br" omitted, but OMITTAG NO was specified.
... 16:16:10 rishida Exp \$</p>
</body>

You may have neglected to close an element, or perhaps you meant to "self-close" an element, that is, ending it with ">" instead of ">".

2. **Info** Line 14 column 93: start tag was here.

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

For this slide we imagine that the W3C Validator is altered slightly so that it tells you how many validation errors are in your file. It will do this using a 'composite message' whose parts are assembled using PHP code as the page is served.

Although we use PHP for these examples, the concepts can be applied to other scripting or coding environments.

Linguistic issues

Handling variables

W3C[®]

W3C[®] QUALITY Assurance

Markup Validation Service v0.7.0

[Home](#) [About...](#) [News](#) [Docs](#) [Help & FAQ](#) [Feedback](#)

Result:

Failed validation

There were 268 validation errors in the file myFirst.html

File:

webdeveloper_localhost_1159293639132.html

Modified:

(undefined)

Server:

Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.8.0.7) Gecko/20080909 Firefox/1.5.0.7

Size:

(undefined)

Jump To:

Results

There were 268 validation errors in the file myFirst.html.

Die Datei myFirst.html enthält 268 Gültigkeitsfehler.

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 12

In the German translation, the order of the two variables may need to be changed.

Linguistic issues
W3C®

Handling variables

```

echo "There were " $.errors. " validation errors in
the file " $. $filename. ". ";

echo "Die Datei " $.errors. " enthält " $. $filename.
"Gültigkeitsfehler. ";

```

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

Typically translators have no access to the actual code, to avoid them introducing bugs into the page. Either the text is extracted or a translation tool masks the code.

Although we are fortunate that we were able to add words after the second variable, due to the English string containing a period, this still didn't produce the right result. The German reads "File 268 contains myFirst.html validation errors."

W3C®

Linguistic issues

Handling variables



Die Datei 268 enthält myFirst.html Gültigkeitsfehler. X

```

echo "There were ".$errors." validation errors in
the file ".$filename.".";

echo "Die Datei ".$errors." enthält ".$filename.
"Gültigkeitsfehler.";
```

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 14

The reason is that the translation process didn't switch the order of the variables.

Linguistic issues
Handling variables

W3C®

W3C®
Quality Assurance

Markup Validation Service v0.7.0

Home About... News Docs Help & FAQ Feedback

Result: Failed validation
There were 268 validation errors in the file myFirst.html

File: webdevelo...localhost_1159293639132.html
Modified: (undefined)
Server: Mozilla/5.0 (Windows NT 5.1; en-US; rv:1.8.0.7) Gecko/20100101 Firefox/1.5.0.7
Size: (undefined)

Jump To:
Results

Die Datei 268 enthält 0 Gültigkeitsfehler.

```

printf(
    "There were %d validation errors in the file %s.",
    $errors, $filename )

printf(
    "Die Datei %s enthält %d Gültigkeitsfehler.",
    $errors, $filename )

```

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 15

So next we try using a printf statement. This has the benefit that text and variable locators all sit within a single string, and the translator can access the items they want to reorder.

Unfortunately, this doesn't help, since PHP still replaces the variables in the string in the order of the variables cited in the following parameters to printf. This causes the 268 to be shown instead of the filename, by converting the integer value to a string. It is unable to find an integer value in the file name, and so presents us with the zero for the number of errors.

Linguistic issues
W3C®

Handling variables



Die Datei myFirst.html enthält 268 Gültigkeitsfehler.

```

printf(
    "There were $errors validation errors in the file
    $filename." )

printf(
    "Die Datei $filename enthält $errors Gültigkeitsfehler."
    )
    
```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 16

By embedding the variable names directly in the printf string, as shown in this slide, we finally achieve the desired result in German.

Nota bene: Successful, or at the very least, cost effective localization in this case is down to the **designer/developer** understanding the potential pitfalls of various approaches to coding. It is not the job of the localization vendor to get this right. It needs to be done as the initial content is created!

You should also be very careful of the assumption that 'This doesn't affect me, since we don't translate the content I develop.' I have seen many, many cases where the thing being developed was later so successful that people wanted to take it to other regions, only to find that they ran into major difficulties because of issues with the translatability of the code or content. It's best to just do it right from the start.

Linguistic issues
W3C®

Handling variables

Die Datei myFirst.html enthält 268 Gültigkeitsfehler.

```

printf(
    "There were %1\%d validation errors in the file %2\%s.",
    $errors, $filename )

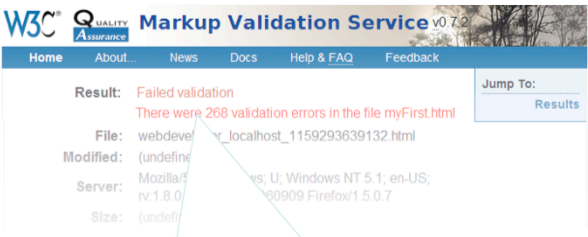
printf(
    "Die Datei %2\%s enthält %1\%d Gültigkeitsfehler.",
    $errors, $filename )

```

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

By the way, there is a way to produce the right effect while using the %d and %s variable markers in a PHP string, but it involves a slightly more complex syntax. This is shown in the above slide. The numeric markers refer to the relevant variable in the parameters that follow the string, even after reordering.

Linguistic issues
Handling variables
W3C



Result: Failed validation
There were 268 validation errors in the file myFirst.html

File: webdev..._localhost_1159293639132.html

Modified: (undefined)

Server: Mozilla/5...; U; Windows NT 5.1; en-US; rv:1.8.0... Firefox/1.5.0.7

Size: (undefined)

Jump To: [Results](#)

1 ошибка

2 ошибки

5 ошибок

При проверке файла myFirst.html была обнаружена 268 ошибок.

```

printf(
    "При проверке файла $filename была обнаружена $errors
    ошибка."
)

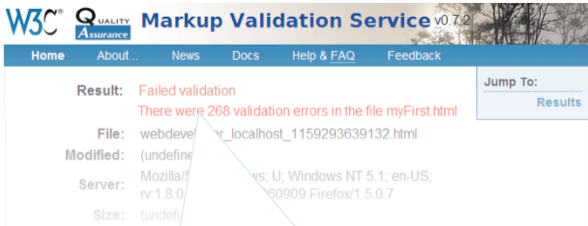
printf(
    "Файл: $filename. Количество ошибок при проверке: $errors."
)
  
```

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

So now we know how to code this type of text in PHP... or do we?

Let's think back to our example of how plurality works in Russian, and we realize that we still have a problem for that language. We only have a single string and it can only be translated one way – yet the Russian requires three variants of the word ошибка, depending on the number that precedes it.

Linguistic issues
Handling variables
W3C



Файл: myFirst.html. Количество ошибок при проверке: 268.

```

printf(
  "При проверке файла $filename была обнаружена $errors
  ошибка."
)

printf(
  "Файл: $filename. Количество ошибок при проверке: $errors."
)

```

1 ошибка

2 ошибки

5 ошибок

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 19

To deal with this, the Russian translator would probably resort to a completely different structure for the text, essentially equivalent to "File: X. Validation errors: Y". This approach requires only one form of ошибка in the invariable string. This is an example of what I call a 'topic-comment' composite message.

Linguistic issues

Sentence-like composite messages

W3C®

The < > is on.


printer stacker

 stapler options


Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 20

So we are beginning to see here that there are two distinct types of composite message. The first is based on a sentence-like approach, and the invariant string can be difficult to translate in some circumstances because of the need for agreement or different word mappings.

In the example above, 'The' should be translated 'el', 'la', or 'las' in Spanish, depending on what word follows it. Also the word 'on' should be translated using three different Spanish words (with different endings).

Linguistic issues

Topic-comment composite messages



Printer: on
Stacker: on
Stapler options: on

Impresora: conectada
Apilador: encendido
Opciones de grapado: activadas

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

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 21

The other approach to designing composite messages is what I like to call the 'topic-comment' approach: you state a topic, then you say something about it.

This approach works much better for the previous slide, since you each comment you associate with a topic can use a different word with the appropriate word endings.

There is a little more to this theory of composite messages than we have mentioned so far, but you can get more information from the W3C Internationalization site at the following URI: <http://www.w3.org/International/articles/composite-messages/> .

Linguistic issues
Text string re-use

W3C®

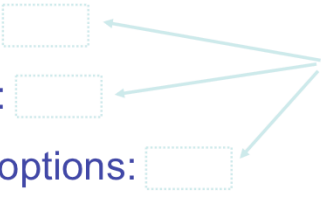
Connectado	-os -a -as
Encendido	-os -a -as
Activado	-os -a -as

Printer:

Stacker:

Stapler options:

On




Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 22

I should, however, mention just one other point. Many designers/developers looking at the English topic-comment arrangement on the previous slide might think to themselves that they could save a little bandwidth by reducing all those instances of the word 'On' to a single string that is used for all comments, ie. they want to re-use strings.

Linguistic issues

Text string re-use



```

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

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

?>

```

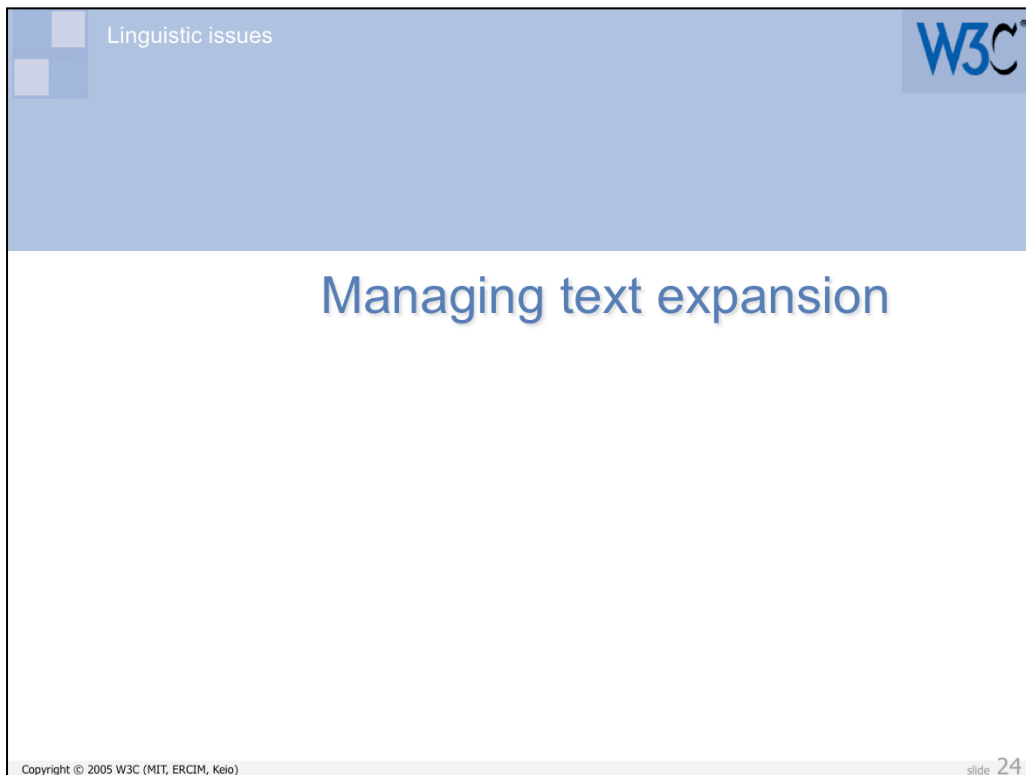
<http://www.w3.org/International/articles/text-reuse/>

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

Tempting as this idea may appear, it will unfortunately introduce insurmountable problems for translation, since the comment is likely to require different agreement forms at the least, and possibly different words altogether, depending on the context.

This slide shows an example of how such a problem may come about by returning the same text from a function for each comment.

Note that I do not want to rule out string re-use altogether – there are situations where it is a sensible approach. But re-use must not occur across different contexts. For more information about this, see the W3C Internationalization article at <http://www.w3.org/International/articles/text-reuse/>.



Linguistic issues

W3C®

Managing text expansion

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

This is a presentation slide. The top header is light blue and contains the text 'Linguistic issues' on the left and the 'W3C' logo on the right. The main body of the slide is white and features the title 'Managing text expansion' in a large, blue, sans-serif font. The bottom of the slide has a thin grey footer bar containing the copyright notice 'Copyright © 2005 W3C (MIT, ERCIM, Keio)' on the left and 'slide 24' on the right.

Now we switch to a very different topic area, that has more to do with the visual layout of the page than the composition of the text.

Linguistic issues

Graphics in fixed width boxes

W3C®

Idioma de la interfaz

☐ arabe

☐ inglés

☐ finlandés

☐ francés

☐ italiano

☐ japonés

☐ ruso

☐ español

☐ sueco

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 25

Let's assume that we want to implement a **fixed-width** box on our page. The text can expand downwards, but not sideways.

Let's also assume that we want a background with a nice gradient behind the title of the box, and that the background has a line across the bottom.

(This slide in Spanish has the title 'Interface Language', and a list of radio buttons to select a language.)

Linguistic issues

Graphics in fixed width boxes

W3C[®]

Bahasa Pengantar
untuk Penelusuran

☐ Arab
☐ Inggris
☐ Finlandia
☐ Perancis
☐ Italia
☐ Jepang
☐ Rusia
☐ Spanyol
☐ Swedia

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 26

As our text expands during translation into Malay, the title occupies two lines. Unfortunately the graphic used for the gradient background is only one line deep, and things now begin to look a mess.

Linguistic issues

Graphics in fixed width boxes

W3C®

Bahasa Pengantar
untuk Penelusuran

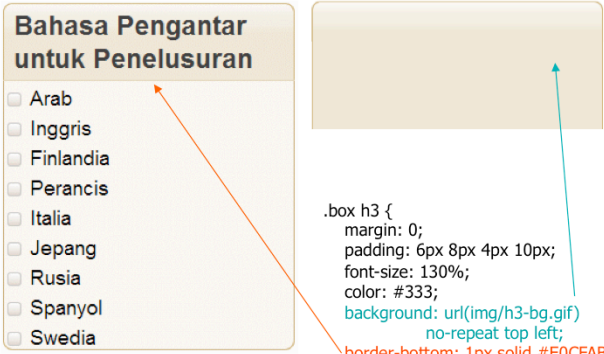
☐ Arab
☐ Inggris
☐ Finlandia
☐ Perancis
☐ Italia
☐ Jepang
☐ Rusia
☐ Spanyol
☐ Swedia

```
.box h3 {  
  margin: 0;  
  padding: 6px 8px 4px 10px;  
  font-size: 130%;  
  color: #333;  
  background: url(img/h3-bg.gif)  
             no-repeat top left;  
  border-bottom: 1px solid #E0CFAB;  
}
```

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

A way to approach this issue is to use a graphic that is three or four lines deep behind the title. By attaching the graphic using the CSS background property, only the amount needed to view the title will actually be shown.

Linguistic issues
Graphics in fixed width boxes
W3C[®]



```
.box h3 {
  margin: 0;
  padding: 6px 8px 4px 10px;
  font-size: 130%;
  color: #333;
  background: url(img/h3-bg.gif)
    no-repeat top left;
  border-bottom: 1px solid #E0CFAB;
}
```

Bulletproof Web Design, Dan Cederholm

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

To get the line to appear in the right place, we simply create it as the bottom border of the heading.

This example uses a technique (and the exact same code and graphic) described in Dan Cederholm's book, *Bulletproof Web Design* (although the text is borrowed from Google's language preferences). This is significant! Dan is not writing about internationalization per se – he is more concerned with people pumping up the text size for accessibility reasons. It just so happens, however, that the same approach helps with localizability.

This is an example that you don't necessarily have to learn new information to deal with internationalization issues – just following existing best practices can be the key in many cases. Note again, however, that we are still talking about the **design and development** of content – not about work that the localizers will do!

Dan's book contains several other recommendations that will benefit internationalization.

Linguistic issues
Language links
W3C[®]

Preferences
[Preferences Help](#) | [About Google](#)

Save your preferences when finished and return to search. Save Preferences

Global Preferences (changes apply to all Google services)

Interface Language Display Google tips and messages in: English
If you do not find your native language in the pulldown above, you can help Google create it through our [Google in Your Language program](#).

Search Language

☒ Search for pages written in any language ([Recommended](#))

☐ Search only for pages written in these language(s):

<input type="checkbox"/> Arabic	<input type="checkbox"/> English	<input type="checkbox"/> Indonesian	<input type="checkbox"/> Romanian
<input type="checkbox"/> Bulgarian	<input type="checkbox"/> Estonian	<input type="checkbox"/> Italian	<input type="checkbox"/> Russian
<input type="checkbox"/> Catalan	<input type="checkbox"/> Finnish	<input type="checkbox"/> Japanese	<input type="checkbox"/> Serbian
<input type="checkbox"/> Chinese (Simplified)	<input type="checkbox"/> French	<input type="checkbox"/> Korean	<input type="checkbox"/> Slovak
<input type="checkbox"/> Chinese (Traditional)	<input type="checkbox"/> German	<input type="checkbox"/> Latvian	<input type="checkbox"/> Slovenian
<input type="checkbox"/> Croatian	<input type="checkbox"/> Greek	<input type="checkbox"/> Lithuanian	<input type="checkbox"/> Spanish
<input type="checkbox"/> Czech	<input type="checkbox"/> Hebrew	<input type="checkbox"/> Norwegian	<input type="checkbox"/> Swedish
<input type="checkbox"/> Danish	<input type="checkbox"/> Hungarian	<input type="checkbox"/> Polish	<input type="checkbox"/> Turkish
<input type="checkbox"/> Dutch	<input type="checkbox"/> Icelandic	<input type="checkbox"/> Portuguese	

SafeSearch Filtering [Google's SafeSearch](#) blocks web pages containing explicit sexual content from appearing in search results.

☐ Use strict filtering (Filter both explicit text and explicit images)
☒ Use moderate filtering (Filter explicit images only - default behavior)
☐ Do not filter my search results.

Number of Results Google's default (10 results) provides the fastest results.
 Display 10 results per page.

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


Note, in passing, an issue related to the Google text I used in the previous example. The dialogue allowed you to select a different language for the user interface from a pull-down list, presumably assuming that your reason for changing was that you couldn't read the current language.

The issue for me is that the names of all the languages are in the language of the current page. Let's assume, for example, that a curious person wanted to see what the interface looked like in Persian, so they selected that language from the list and clicked on the 'Save Preferences' button.

Linguistic issues

Language links





The screenshot shows the Persian version of the Google homepage. At the top, it says 'تنظیمات' (Settings) and 'Google'. Below the search bar, there are links for 'تنظیمات' and 'تنظیمات را ذخیره کنید'. The main content area has a dropdown menu for 'زبان نمایش فرکانس' (Frequency display language) set to 'فارسی' (Persian). Below this is a section for 'زبان جستجو' (Search language) with a list of languages including English, Spanish, French, German, Italian, Japanese, Korean, Russian, Chinese, Hindi, Bengali, Thai, Vietnamese, Indonesian, Malay, and others. At the bottom, there is a section for 'شمار یافته ها' (Number of results) and 'پنجره یافته ها' (Result window).

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 30

Assuming that they would be able to find their way back to the appropriate dialogue box to get back to English (which would require them to remember which link to hit on the thankfully uncluttered Persian Google home page), that they can remember which is the required select list, and that they can do so in spite of the mirror-imaging of the page when using Arabic script, they would then be faced with what you see on the next slide.

Linguistic issues

Language links





Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 31

Note that the names of languages are all in Persian, and are sorted by Persian rules.

Which selection would get you back to English ??

(Hint: if you want to explore like this, use a different tab or window for your explorations, and leave the original dialogue available in another for when you want to reset to your current language.)

Of course, the point is really that a Persian person taken to the English site may have as much trouble finding their way to the appropriate user interface language as the curious explorer does in getting back. In my opinion it would help a great deal to write each language name in its own script and language. You can read more about that in the W3C Internationalization article at <http://www.w3.org/International/questions/qa-navigation-select>.

Linguistic issues

Form field label positioning

W3C[®]

Global settings

Interface language	English
Search language	English
Number of results	10
Save preferences	

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 32

Lets take a moment to explore another potential issue related to the length of text in translation.

Let's continue to assume a situation where text appears in a fixed width box. We will apply the same approach we discussed earlier to deal with the title of the box. The issue this time will be that we have used a table to apply form labels to the left side of the form entry field to which they apply. Our initial source text is in English.

Linguistic issues
W3C[®]

Form field label positioning

Global settings

Interface language English

Search language English

Number of results 10

Acuan Umum

Bahasa Pengantar di Antar Muka Inggris

Bahasa Pengantar untuk Penelusuran Inggris

Jumlah Hasil Penelusuran 10

Allgemeine Voreinstellungen

Sprache der Benutzeroberfläche Englisch

Suchsprache Englisch

Anzahl der Ergebnisse 10

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 33

The English looks nice enough. The Malay, on the other hand, looks pretty ugly. The large expansion factor produces unfortunate stacking of the text on the left, and large white spaces to the right. Although the box expands vertically to hold all the text, we are wasting a lot of space and decreasing the amount of information that will appear in the reader's initial screen (you can imagine that this would be compounded by other fixed width boxes on the page).

With the German translation we have a different problem. The long word Benutzeroberfläche doesn't wrap, and so pushes the select boxes beyond the width of the fixed box container. This has the potential to badly affect the layout of other parts of the screen.

Linguistic issues

Form field label positioning

W3C[®]

Global settings

Interface language
English

Search language
English

Number of results 10

Save preferences

Acuan Umum

Bahasa Pengantar di Antar Muka
Ingriss

Bahasa Pengantar untuk Penelusuran
Ingriss

Jumlah Hasil Penelusuran 10

Simpan Acuan

Allgemeine Voreinstellungen

Sprache der Benutzeroberfläche
Englisch

Suchsprache
Englisch

Anzahl der Ergebnisse 10

Einstellungen speichern


Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 34

You may want to consider avoiding table cells in such constrained circumstances. This slide shows how the text would look if the input fields were just in a paragraph with the label text.

All the boxes now look fine, and although there is a very slight increase in vertical height overall, we have removed the problems seen with the Malay and German text on the previous slide.

Let's note, again, that this is down to the way the page is **designed/developed**, not the way it is localized. That's a fundamental message of this presentation. Internationalization during design and development removes significant barriers to deploying your content globally.



Separating content & presentation

- Linguistic issues
- Separating content & presentation
- Cultural issues
- What this means to me

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

Now we are going to look at the benefits to localization of another good design/development best practice that you would hopefully adopt anyway: the separation of content, presentation and behaviour.

Separating content & presentation
The basics
W3C®


Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 36


People at this conference should be familiar with the idea that content and presentation should be kept separate, if you want manageable and easily maintainable web sites.

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
Style issues for Japanese




- ◆ problems of resolution to support bold and italics in small characters on-screen => different ways of emphasizing text

これは日本語です。

これは日本語です。


Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 37

Here are some ways in which typographic differences may appear between language versions of the same content. It is much easier to apply each of these typographic differences if you can do so via a CSS style sheet, rather than searching through the HTML or script code.



Separating content & presentation

Style issues for Japanese




- ◆ problems of resolution to support bold and italics in small characters on-screen => different ways of emphasizing text
- ◆ different fonts, and possibly font-size, need to be used for font-family assignments
- ◆ line heights, paragraph spacing may need to be adapted, since characters may be more complicated

Leading the Web to
its full potential...


การนำไปสู่ Web ที่เต็ม
ไปด้วยศักยภาพ

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 38

 Separating content & presentation

Style issues for Japanese



- ◆ problems of resolution to support bold and italics in small characters on-screen => different ways of emphasizing text
- ◆ different fonts, and possibly font-size, need to be used for font-family assignments
- ◆ line heights, paragraph spacing may need to be adapted, since characters may be more complicated
- ◆ wrapping rules may need adjustment (eg. there are no spaces between words in Chinese)
- ◆ no upper- vs. lower-case distinction
- ◆ no convention of distinguishing between proportional and mono-spaced fonts

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 39

[illegible]

You should also consider separation of content and presentation when adding scripting. Let's suppose that we wanted to load some JavaScript after this basic test page has loaded which would automatically add a list of tests on the page to the top right corner. (We may actually want to add links to these tests, but I have resisted that temptation so that the following slides will contain the code examples.)



Separating content & presentation
Don't forget your scripting



```
function makeList () {  
  var headings = document.getElementsByTagName('h2');  
  var div = document.createElement('div');  
  div.style.cssFloat = 'right';  
  div.style.styleFloat = 'right';  
  div.style.width = '20%';  
  div.style.padding = '1em';  
  div.style.border = '1px solid teal';  
  div.style.backgroundColor = '#FFCC99';  
  div.style.fontSize = '90%';  
  div.style.margin = '1em';  
  div.style.color = '#996633';  
  var title = div.appendChild(document.createElement('div'));  
  title.style.fontWeight = 'bold';  
  title.appendChild(document.createTextNode('Tests on this page:'));  
  for (i=0;i<headings.length;i++) {  
    var test = div.appendChild(document.createElement('p'));  
    test.appendChild(document.createTextNode(headings[i].childNodes[0].data));  
  }  
  body = document.getElementsByTagName('body')[0];  
  body.insertBefore( div, body.firstChild );  
}
```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 41

Here is a simple function that could be used to add the required text. It creates a div, gets a list of level two headings, and adds the text of the headings to the list.

Separating content & presentation

Don't forget your scripting

W3C[®]

```
function makeList () {
  var headings = document.getElementsByTagName('h2');
  var div = document.createElement('div');
  div.style.cssFloat = 'right';
  div.style.styleFloat = 'right';
  div.style.width = '20%';
  div.style.padding = '1em';
  div.style.border = '1px solid teal';
  div.style.backgroundColor = '#FFCC99';
  div.style.fontSize = '90%';
  div.style.margin = '1em';
  div.style.color = '#996633';
  var title = div.appendChild(document.createElement('div'));
  title.style.fontWeight = 'bold';
  title.appendChild(document.createTextNode('Tests on this page:'));
  for (i=0;i<headings.length;i++) {
    var test = div.appendChild(document.createElement('p'));
    test.appendChild(document.createTextNode(headings[i].childNodes[0].data));
  }
  body = document.getElementsByTagName('body')[0];
  body.insertBefore( div, body.firstChild );
}
```

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 42

Note how we are adding style information directly to the DOM while running this script. This is really obvious in this example, since there is such a lot of it. It is particularly tempting to do this sort of thing if you just want to add a single style effect, such as bolding, to text.

Separating content & presentation
W3C

Don't forget your scripting

```
function makeList () {
  var headings = document.getElementsByTagName('h2');
  var div = document.createElement('div');
  div.setAttribute( 'id', 'testList' );
  var title = div.appendChild(document.createElement('div'));
  title.appendChild(document.createTextNode('Tests on this page:'));
  for (i=0;i<headings.length;i++) {
    var test = div.appendChild(document.createElement('p'));
    test.appendChild(document.createTextNode(headings[i].childNodes[0].data));
  }
  body = document.getElementsByTagName('body')[0];
  body.insertBefore( div, body.firstChild );
}
```

```
div#testList {
  float: right;
  width: 20%;
  padding: 1em;
  border: 1px solid teal;
  background-color: #FFCC99;
  font-size: 90%;
  margin: 1em;
  color: #996633;
}

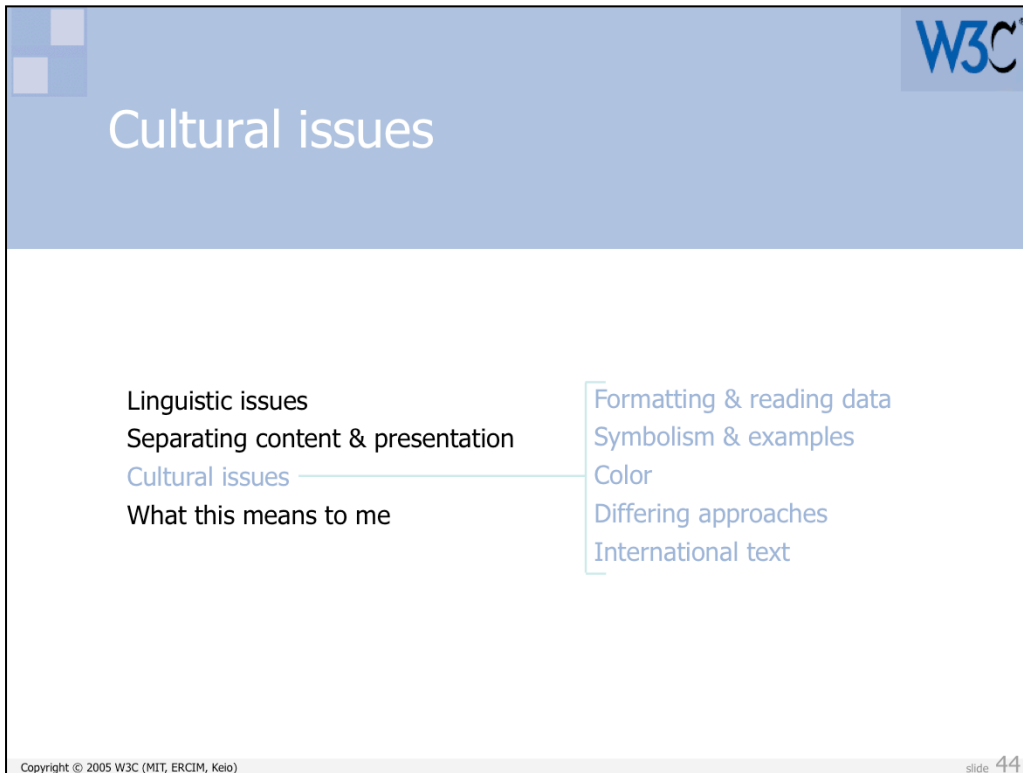
div#testList div {
  font-weight: bold;
}
```

see also: [Dom Scripting, Jeremy Keith](#)
slide 43

Copyright © 2005 W3C (MIT, ERCIM, Keio)

This version of the same function shows a much better approach. We assign an id attribute to the box, then move all the styling information to a CSS file, referencing the markup via the id. This makes the code much cleaner and makes it easier to manage the styling.

Again, this technique is recommended as a standard best practice in Jeremy Keith's book *Dom Scripting* (which contains many other useful ideas along similar lines). It is another good example of how good web design benefits localization.



The slide features a blue header with the title 'Cultural issues' and the W3C logo. Below the header, a list of topics is presented in two columns. A horizontal line connects 'Cultural issues' in the left column to the right column. The footer contains copyright information and the slide number 44.

Cultural issues


W3C®

- Linguistic issues
- Separating content & presentation
- Cultural issues
- What this means to me


- Formatting & reading data
- Symbolism & examples
- Color
- Differing approaches
- International text

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

This section will look at a very different set of issues – those related to cultural differences in design.



Cultural issues



Formatting & reading data

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 45

Cultural issues

Formatting & reading data

W3C[®]

03/02/01
2.3.01
2001-III-2
2001年3月2日

Date*:

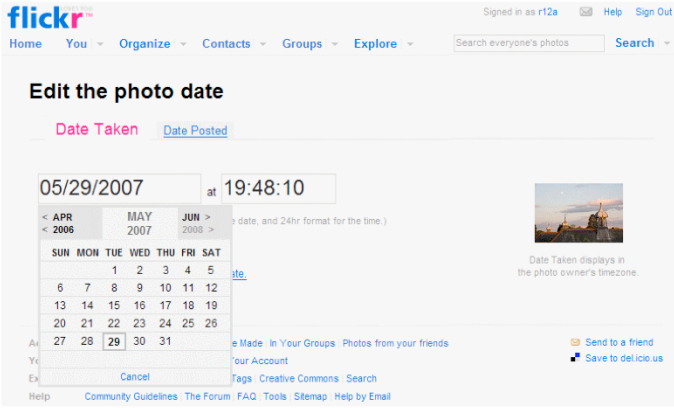
Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 46

The date at the top of the slide is ambiguous in three ways. This is a bad way to generate dates for a page – it is better to use a name for the month (or, in some specialized cases, you may be able to use a four-digit year and the order year, month then day).

Note also how the expected separators, leading zeros, etc vary from culture to culture.

The biggest issue, however, is not displaying the date correctly, but recognizing a date supplied by a user if you haven't adequately signposted the order or format in which you expect to receive data. You need to make sure that your expectations are clear if you want to stand any chance of recognizing what the user is typing in.

Cultural issues
W3C[®]
Formatting & reading data



Date*:

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

Often, using a graphical calendar can provide a more user friendly and reliable method for users to indicate dates. (Be careful to ensure that your calendar allows enough space for translations of the month and day abbreviations in other languages, of course!)

Bear in mind, also, that people in some parts of the world use local calendars.

Cultural issues
Formatting & reading data

W3C®

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

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

When Chinese people write their name they normally write the family name first and given name last. A form that asks a Chinese user to enter their first and last names can be very confusing for them. Better say 'family' and 'given' name.

The Malay person above has only one name, Isa. 'bin' means 'son of', and Aman is his father's name.


A similar situation applies to the person from southern India whose name appears at the bottom.

There are other ways in which names can vary, including double family names for Spanish people, and patronymics for Russians.

When creating forms for names, ask yourself what you will do with the name. If you won't process it at all, allow people to enter their whole name as they would usually write it. However, if you are expecting, for example, to use part of the name to address people, you may find that you can't simply work out what to call people working from Anglo-Saxon expectations of how names are used. You may need a special field that asks how the user likes to be addressed. Also be careful about choosing a part of the name for sorting – people sort names in very different ways around the world.

Cultural issues

Formatting & reading data



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

Country: United States

First name:

Last name:

Address:

City:

State: AZ

Zip code:

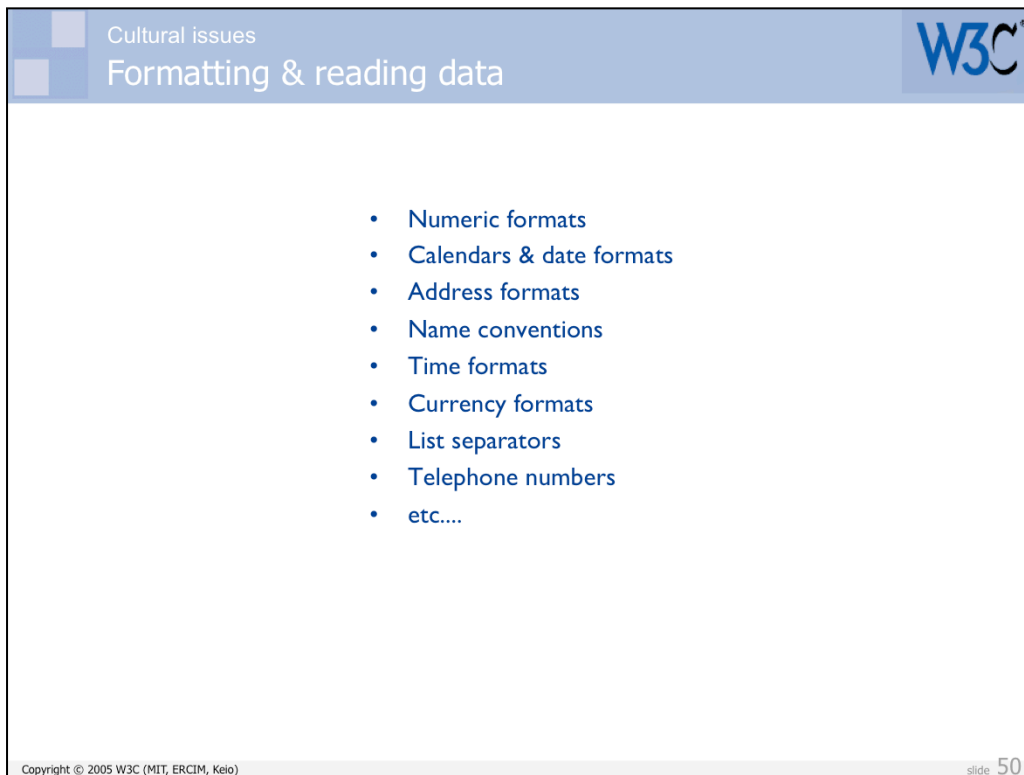
Telephone: ()

Application date:

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


You also need to consider that addresses look quite different from country to country. Russian and Japanese addresses are written from the general to the specific, top to bottom. You may need to figure out how to produce these different orderings for forms.

Also, the name of the Russian person above is in the dative case (expressing the idea of 'to the person'). If you asked her to simply supply her name in a separate box, she would probably write Юлия Селиванова, rather than Юлии Селивановой. How will you deal with that?



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 50

There are a number of ways in which formats differ around the world. Note that recognizing information input into general forms can be more difficult than producing form templates in the right way.



Cultural issues




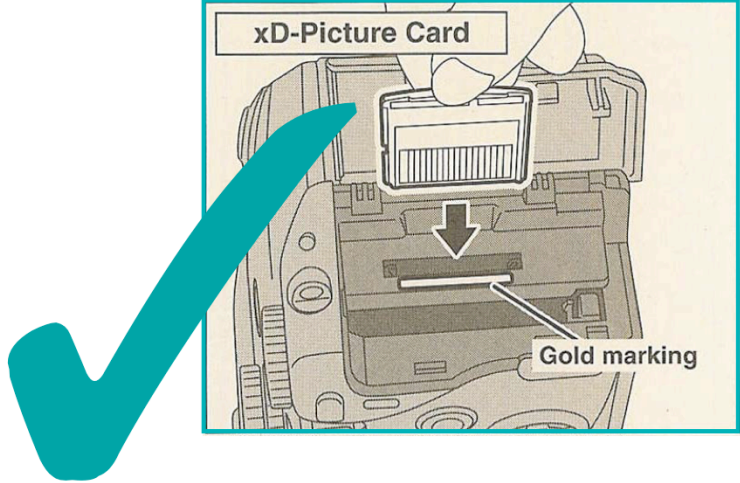
Symbolism & examples

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 51

Cultural issues

Symbolism & examples





Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 52

This check symbol means 'correct' or 'ok' in many countries. In some countries, however, such as Japan, it can indicate 'incorrect'. Japanese often convert check marks to circles (their symbol for 'correct') as part of the localization process.




Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 53

The circles in the columns of this board indicate that space is available, not that there are 0 seats left. It is the equivalent of the check mark.

Cultural issues
Symbolism & examples

W3C®



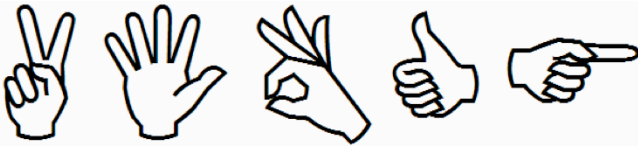
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 54

This illustration of sports items is not representative of sports played in the UK, and may need to be changed.

Cultural issues

Symbolism & examples

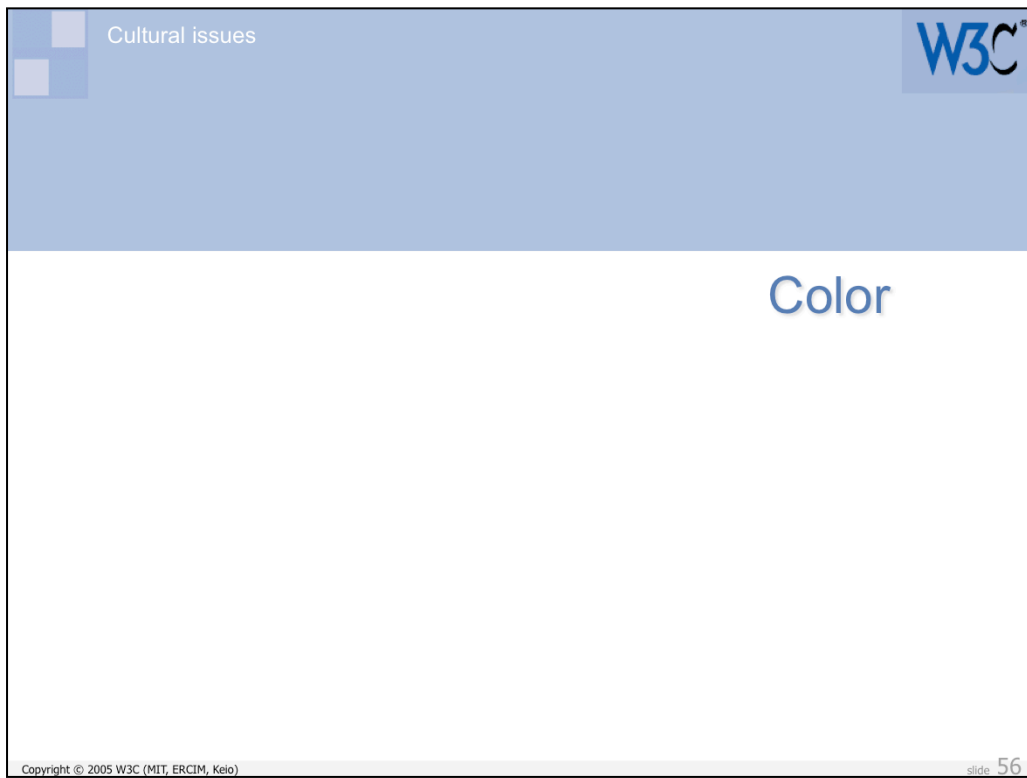
W3C[®]



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 55

Gestures and sometimes body language can often give completely the wrong message, and should be used with extreme care.



Cultural issues

W3C®

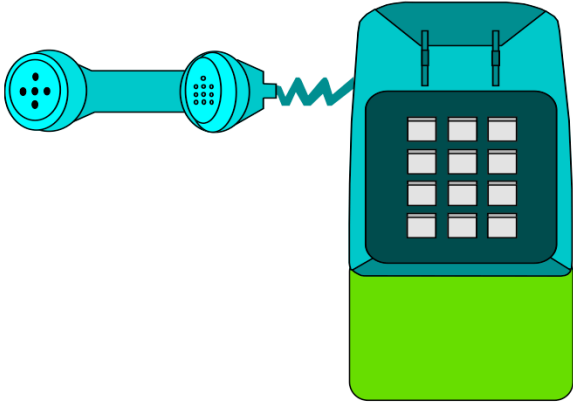
Color

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 56

This slide features a blue header bar with the text "Cultural issues" on the left and the W3C logo on the right. The main content area is white and contains the word "Color" in blue. The footer is a light gray bar with copyright information and the slide number "56".

Cultural issues
Color

W3C®



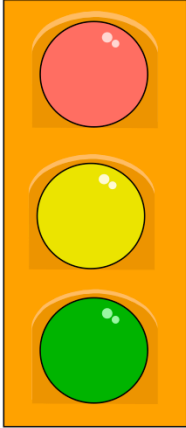
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 57

This phone is likely to be perceived immediately as a public telephone in Japan, due to the conventional use of the green color there. In most other parts of the world, this cue is missing. So colors have conventional roles that differ from culture to culture.

Cultural issues

Color


W3C®




Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 58

Color names also differ from culture to culture, dependent on context. British people often call the middle light here amber, whereas Americans call it yellow. Japanese speaking English will often refer to the bottom light as blue.

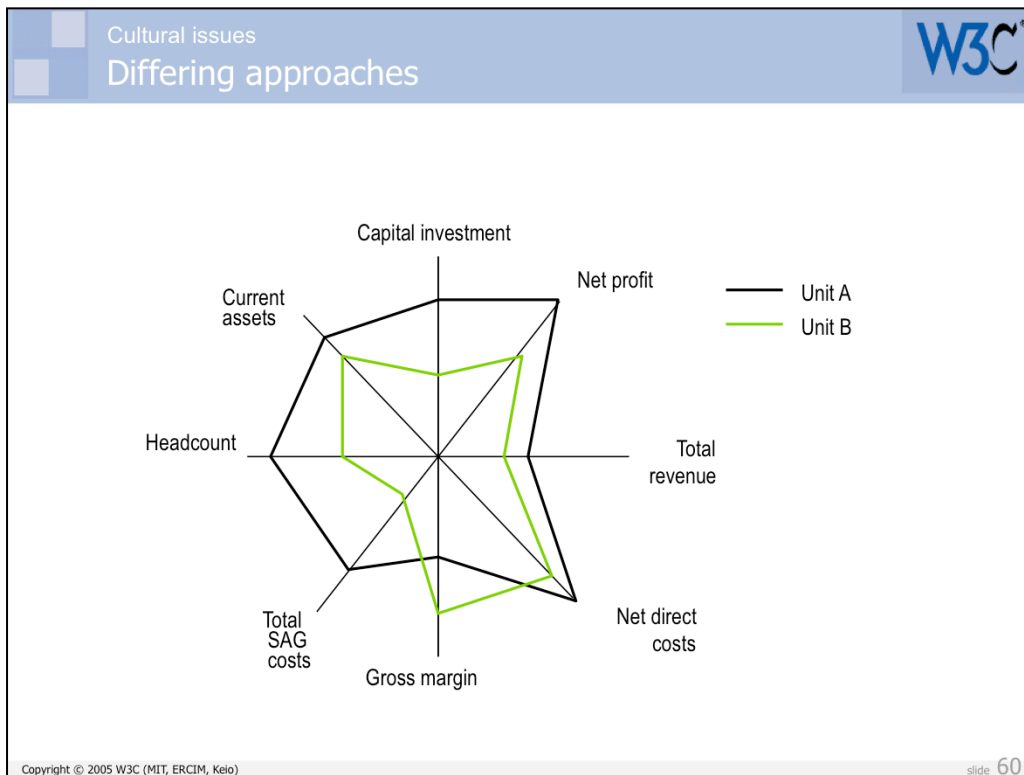


Cultural issues

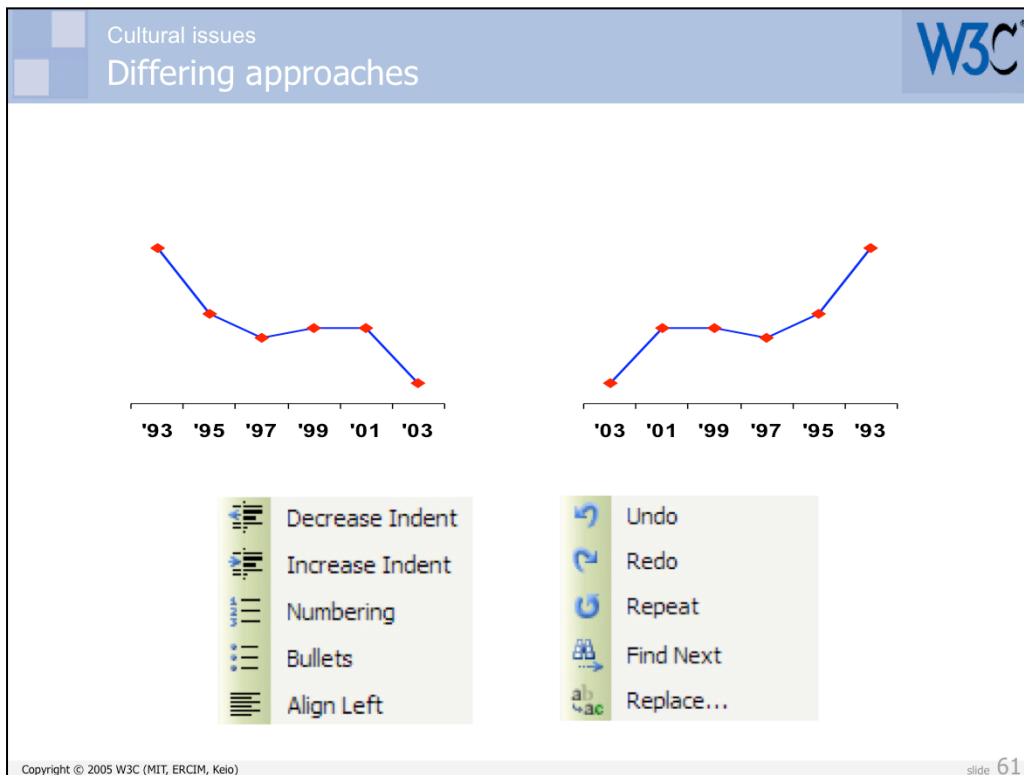


Differing approaches

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 59




People do things in different ways in different parts of the world. For example, Lotus 1-2-3 was relaunched in Japan with the radar chart after it was discovered that this was a very common way of representing comparative data there.



In the Middle East, you may find that tables, spreadsheets, collated pictures and the like need to flow right to left, rather than left to right.

Some graphics with directional bias may need to be mirrored or changed for a predominantly right-to-left context.

W3C[®]

Cultural issues
Differing approaches


"... 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 62

Then there are more fundamental issues about whether the application, product or solution you are developing will actually fit into the foreign culture at all.

Cultural issues
Differing approaches
W3C®



Web Site Directory Sites organised by subject.

<p>Arts & Humanities Literature, History, Photography...</p> <p>Business & Economy B2B, Shopping, Investments, Property...</p> <p>Computers & Internet Internet, Reviews, Software, Games...</p> <p>Education UK, Ireland, Universities...</p> <p>Entertainment Humour, Movies, Music, Actors...</p> <p>Government UK, Ireland, Politics, Law...</p> <p>Health Medicine, Drugs, Diseases, Fitness...</p>	<p>News & Media Newspapers, Weather, TV...</p> <p>Recreation & Sport Sport, Hobbies, Travel, Motoring...</p> <p>Reference Maps, Dictionaries, Phone Numbers...</p> <p>Regional UK, Ireland, Countries...</p> <p>Science Animals, Geography, Engineering...</p> <p>Social Science Economics, Languages, Psychology...</p> <p>Society & Culture People, Food & Drink, Environment, Sexuality...</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 63

The following slides show how Yahoo **adapts the content** on its various local home pages, rather than just translating it. This may be something you also need to consider.

Cultural issues
W3C®

Differing approaches

Guide Web - Classement thématique de sites Web
[Suggérer un site](#)

[Actualités et médias](#)
[Journaux](#), [Télévision](#), [Météo](#)...

[Commerce et économie](#)
[B2B](#), [Shopping](#), [Emploi](#),
[Immobilier](#)...

[Informatique et Internet](#)
[Internet](#), [Logiciels](#), [Fonds d'écran](#)...

[Santé](#)
[Diététique](#), [Médecine](#), [Thalasso](#)...

[Enseignement et formation](#)
[Primaire](#), [Secondaire](#), [Supérieur](#)...

[Institutions et politique](#)
[Ministères](#), [Droit](#), [Politique](#)...

[Sciences et technologies](#)
[Animaux](#), [Astronomie](#), [Physique](#)...

[Sports et loisirs](#)
[Foot](#), [Tourisme](#), [Auto/Moto](#), [Jeux](#)...

[Art et culture](#)
[Littérature](#), [Cinéma](#), [Musique](#), [BD](#)...

[Divertissement](#)
[Tests/Quiz](#), [Loteries](#), [Humour](#),
[Sorties](#)...

[Classement géographique](#)
[Pays](#), [Europe](#), [France](#), [Paris](#)...

[Références et annuaires](#)
[Dictionnaires](#), [Annuaire](#),
[Cartes/Atlas](#)...

[Société](#)
[Enfants](#), [Gastronomie](#),
[Rencontres](#)...

[Sciences humaines](#)
[Archéologie](#), [Histoire](#), [Psychologie](#)...

THUS: Personalized according to 2004?

See also:

Yahoo! Mail - FREE!

1 - 5000 US

in ready logs

r more addresses

D in Riyadh

own-assaulted in

r auction

of public-service

on sex in

only - Market

24% heading 1000

France - weather

id's mobile!

Use

- Jobs - Property

on? Head over to not get the format loads to help you.

news

in 15% online

re to get a quote

ions League

s in tatters. Read have to say about

ees - Games - TV

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 64

Cultural issues
W3C®

Differing approaches




The screenshot shows the Yahoo! Japan homepage with various category links. An orange arrow points to the 'Arts and Culture' (芸術と人文) category, which includes links for 'Photography, Architecture, Museums, History, Literature' (写真, 建築, 美術館, 歴史, 文学).

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 65

Cultural issues

Differing approaches



Directory siti Web - I siti ordinati per argomento.

[Proponi il tuo sito](#)

[Economia e aziende](#)
[Lavoro](#), [Finanza](#), [Shopping](#), [B2B...](#)

[Aree geografiche](#)
[Paesi](#), [Città italiane](#), [Meteo](#), [Mappe...](#)

[Attualità e media](#)
[Giornali](#), [TV](#), [Gossip...](#)

[Arte e cultura](#)
[Letteratura](#), [Racconti erotici](#), [Moda...](#)

[Scienza e tecnologia](#)
[Cellulari](#), [Animali](#), [Ecologia...](#)

[Politica e istituzioni](#)
[Governi](#), [Diritto](#), [Fisco](#), [Partiti...](#)

[Risorse](#)
[Calendari](#), [Elenchi telefonici...](#)

[Società e culture](#)
[Sesso](#), [Amore](#), [A tavola](#), [Oroscopo...](#)

[Divertimento e spettacolo](#)
[Musica](#), [Cinema](#), [Celebrità](#), [Fumetti...](#)


[Informatica e internet](#)
[Giochi](#), [Cartoline](#), [Chat](#), [Software...](#)

[Sport e tempo libero](#)
[Viaggi](#), [Sport](#), [Calcio](#), [Auto...](#)

[Medicina e salute](#)
[Gravidanza](#), [Benessere](#), [Malattie...](#)

[Scienze umane e sociali](#)
[Sessuologia](#), [Storia](#), [Geografia](#), [Blog...](#)

[Istruzione e formazione](#)
[Università](#), [Scuole](#), [Dizionari](#), [Libri...](#)



The screenshot shows the Yahoo! homepage with various sections: 'In the News' featuring headlines about UK troops, anti-terror police, and a car bomb; 'FTSE Closed' with market data; 'Today on Yahoo!' with links to Games Domain, Home Insurance, and Chelsea's Champions League; and a 'Personal Assistant' section at the top right.

Copyright © 2005 W3C (M)


slide 66



Cultural issues

International text

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 67

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 68

Even if you expect readers to use English on your site, be careful how you write your English. Don't expect people who use English as a second language to understand all the idioms, words or concepts you are familiar with. (This slide is just an example to make the point that English people may not even understand American texts – see the bits in blue.)

Also consider the complexity of your grammar. Short, simple sentences can often help non-native speakers.



Cultural issues
International text



Montre qui tu es !


The 5 R's

Nothing sucks like an Electrolux!


Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 69

Of course, look out for difficult situations when it comes to translation.



Cultural issues
Visual puns





Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 70

Be wary also of visual puns – ie. graphics that rely on the user speaking a language to be understood.

Cultural issues
Visual puns

W3C®


HangUp

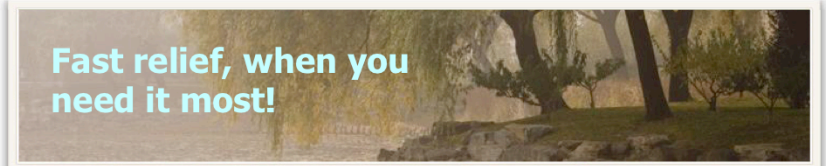
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 71

For example, the French and Japanese translations of ‘Hang up’ have nothing to do with ropes or things hanging. If you translate the icon label, a French or Japanese person will most likely be confused by the choice of graphic!

Cultural issues

International text

W3C[®]




Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 72

If you wanted to translate this text to Russian, and you were supplied with a jpeg file, you'd have to carefully rub out the English, then redraw the complicated background, before you could finally add the Russian text. That would take a huge amount of time. Alternatively, if you were provided with a layered file, that puts text on one layer and background on another, it would be very quick and easy to produce the translation. Think, therefore, about how you go about the process of handing things off for localization.


Of course, an even better approach would be to use CSS positioning for the text. That would make it searchable and selectable. Try to avoid using readable pixellation when you can.




What this means to me

- Linguistic issues
- Separating content & presentation
- Cultural issues
- What this means to me

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 73

What this means to me

Some lessons from this presentation




- Use topic-comment arrangements for composite messages, if you can, rather than sentential arrangements
- Use coding approaches that allow syntactic flexibility when creating composite messages
- Be careful about reuse of text strings – ensure that the context of reuse is always the same
- Think about how to cope with text that expands in translation – for example, use larger background images and avoid placing labels to the side of form entry fields, if you can
- Separate content from presentation – including in JavaScript (use id or class names)
- Don't assume that people format dates, addresses, names, etc like you. Allow for alternative approaches, where appropriate.
- Ensure that you indicate clearly how users should enter information such as dates – in fact, consider graphical approaches that allow you to pick a date from a pop-up.
- Think carefully about how you intend to break up and reuse things like names and addresses, and the implications of that for multicultural formats.
- Be prepared to change graphics if they represent a particular culture or set of cultural aspirations. Also be careful about assuming common understanding of symbols, color, and ways of doing things. Design in a way that allows for adaptation, where necessary, and use in-country knowledge to check that things work as you expected.
- Be careful about how you write your text if you expect non-native speakers of English to read your content.
- Remember that graphics may need to be translated – not just text.
- Keep text out of graphics where you can. Where you can't, supply layered files for translation.
- Check out whether you are incorrectly assuming that people in other parts of the world see their world and do things in the same way as you.

Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 74


This slide summarizes some of the practical takeaways from this presentation.

The presentation is not designed to give you a thorough overview of potential internationalization and localization issues – we would need longer for that. It aims to provide you with a few practical takeaways, but more importantly it aims to get you thinking about what internationalization is all about – to take you out of your comfort zone, and help you realize that if you want your content to wow people outside your own culture and language, you need to build in certain flexibilities and adopt certain approaches during the design and development – not as an afterthought. Otherwise you are likely to be creating substantial barriers for worldwide use.

The presentation also aims to show that, although using Unicode is an extremely good start to making your stuff world-ready, using a Unicode encoding such as UTF-8 throughout your content, scripts and databases is only a start. You need to worry about whether translators will be able to adapt your stuff linguistically, but you also need to also consider whether graphics and design are going to be culturally appropriate or can be adapted, and whether your approaches and methodologies fit with those of your target users.



What this means to me

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 – don't treat internationalization as an afterthought!
- not always doing extra work – maybe just working in a better way


Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 75

What this means to me
Recommendations


- be aware that well localized deliverables and content begins with **you**, not the localization folks
- find out about good internationalization practice

Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 76

Remember also that, even though you think you don't deal with content that will be internationalized now, you may well need to in the future.

What this means to me
W3C can help


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



Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 77

The W3C is trying to provide useful advice at <http://www.w3.org/International/>

We could always do with help and support for this.

Getting help

Topic & techniques indexes

W3C®

W3C HTML Topic Index - World Wide Web

[Language](#)

- lang
- Content-Language
- browser settings
- declaring
- flags
- Hans and Hant language codes
- hreflang
- IANA language tags
- ISO language codes
 - Link: ISO 639 Codes for the Representation of Names of Languages
 - 2-letter and 3-letter codes
 - Action: 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
- xmllang

[Markup & text](#)

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

W3C HTML Techniques Language Tagging

[Choosing language values](#)

- How to choose language values
 - W3C techniques document (Authoring Techniques for XHTML & HTML, Internationalization: Specifying the language of content)
 - Language tags in HTML 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
 - UA issues: none
- Where possible, use the codes *en-GB* and *en-US* to refer to Simplified and Traditional Chinese, respectively
 - UA issues for ?

[Particularly useful links](#)

- IANA Assigned Language Tags
 - IANA's 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 8.1)
- Language identification
 - xml:lang in the XML spec (section 2.12)

[Test data](#)

- Automatic font assignment for CJK text
 - W3C test page
- Automatic font assignment for CJK text
 - W3C test results

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 78

What this means to me

Remember:
Portability, scalability, interoperability + localizability !

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

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 79