



Designing International Web Pages: Some Practical Tips

Richard Ishida
W3C Internationalization Activity Lead



This presentation will look at three of the many aspects of designing internationalized web pages, and attempt (without, by any means, covering the topic exhaustively) to produce a number of guidelines for designers and developers.

The W3C is trying to provide useful advice at <http://www.w3.org/International/> and you will find there articles that take these ideas further. Some of these will be pointed to from the upcoming slides.

We could always do with help and support for this outreach work.

Text size

W3C


Text size

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 3

W3C
Text size
Text size in translation

DochuLa Pass



Around 3,100 m high, in Bhutan.

© This photo is public. [Change?](#)

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

953 views / [comments](#)

Language	Translation	Ratio
Korean	조경	0.8
English	views	1
Chinese	次檢視	1.2
Portuguese	visualizações	2.6
French	consultations	2.6
German	-mal angesehen	2.8
Italian	visualizzazioni	3

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

This slide shows an example of how text can expand (and shrink) when translated (here from English).

Text size
Compound nouns

W3C

Input processing features

Eingabeverarbeitungsfunktionen


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

One potential issue here is that some languages have very long words, which may not fit within the specified max-width.


Text size
Text size in graphics

W3C

Seals and sealions at Pier 39



Seals and sealions at Pier 39



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


If you are using graphic text, beware! Not only is it more difficult to translate graphic text, but graphic text is often squeezed into small spaces, making it difficult to replicate the source in another language. Here, the Portuguese text doesn't fit in the available horizontal space.

Note, also, that complex characters found in languages like Chinese, Japanese, Korean, and many other scripts would be very difficult to read in such small sizes.


Text size
Text size in graphics

W3C

Seals and sealions at Pier 39



Seals and sealions at Pier 39
Encargar impresión



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

Here Flickr has resorted to using just icons with mouse-overs in the translated UIs to get around the expansion issues they would have otherwise had.

Text size

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 8

Here is 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.

Text size

Form field label positioning

W3C

The image shows three side-by-side screenshots of web settings panels. The first panel, titled 'Global settings', has labels on the left and values on the right. The second panel, titled 'Acuan Umum', has labels on the left and values on the right, with some labels wrapped. The third panel, titled 'Allgemeine Voreinstellungen', has labels on the left and values on the right, with the label 'Sprache der Benutzeroberfläche' being very long and not wrapping.

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 9

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.

Text size

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
Ingriiss

Bahasa Pengantar untuk Penelusuran
Ingriiss

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 10

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.

The slide features a blue header with the text 'Text size' and 'Graphics in fixed width boxes' on the left, and the W3C logo on the right. The main content is a yellow box with a gradient background and a horizontal line at the bottom. The box is titled 'Idioma de la interfaz' and contains a list of radio buttons for language selection: arabe, inglés, finlandés, francés, italiano, japonés, ruso, español, and sueco. The slide number '11' is in the bottom right corner, and the copyright notice 'Copyright © 2005 W3C (MIT, ERCIM, Keio)' is in the bottom left corner.

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.)

Text size

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 12

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.

W3C
Text size

Graphics in fixed width boxes

**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 13

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.

Text size
W3C

Graphics in fixed width boxes

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;
}

```

Bulletproof Web Design, Dan Cederholm

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

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.



When creating tabs, don't draw the background as a single graphic or a collection of graphics tailored to fit the source text. This will break both internationalization and accessibility.



Instead use something like Doug Bowman's 'sliding doors' technique (<http://alistapart.com/articles/slidingdoors/>). This allows the graphic part of the tab to expand with the text.



Note, again, that this also aids accessibility. The slide shows the result of increasing text size on the tabs.

One other thing to note, however, is that you should leave some space for expansion to the right of the tabs (or left in a right-to-left script). You may otherwise need to design the CSS so that the tabs can be wrapped and still look good.

Text size
Round cornered boxes

W3C

The value of internationalization

Internationalization significantly affects the ease of the product's localization. Retrofitting a linguistically- and culturally-centered deliverable for a global market is obviously much more difficult and time-consuming than designing a deliverable with the intent of presenting it globally. (Think back to the Y2K effort and trying to "shrink" two-character year fields that were built on the assumption of "1900").

So clearly, internationalization occurs as a fundamental step in the design and development process, rather than as an afterthought that can often involve awkward and expensive re-engineering.

further reading

- W3C i18n resources relating to internationalization


Author: Richard Ishida, W3C; Susan K. Miller, Boeing

Content first published 2005-12-05. Last substantive update 2005-12-05 GMT. This version 2006-10-19 19:25 GMT

For the history of document changes, search for `cp-i18n` in the i18n blog.

Copyright © 2005-2006 W3C[®] (MIT, ERCIM, Keio). All Rights Reserved. W3C liability, trademark, document use and software licensing rules apply. Your interactions with this site are in accordance with our [public](#) and [i18n](#) privacy statements.

Tell us what you think @English
Send us a comment
Subscribe to an RSS feed
View resources
Hosted page views



In the W3C's i18n articles we use a box with rounded corners.

Text size

Round cornered boxes

W3C

The slide displays five examples of web site navigation boxes, each with rounded corners and a light background. The boxes are arranged in a grid-like fashion. Each box contains text in a different language and includes icons for various actions like sending comments, subscribing to RSS feeds, and viewing home page news. The languages shown are English, Polish, Arabic, Russian, and Chinese.

Tell us what you think (English)
Send us a comment
Subscribe to an RSS feed
New resources
Home page news.

Powiedz nam co myślisz (po angielsku)
Wyślij komentarz
Prenumeruj kanał RSS
Nowe źródła
Wiadomość ze strony głównej.

قل لنا رأيك (بالانجليزية).
أبعث تعليق.
الإعتماد إلى خدمة RSS.
موارد جديدة.
الصفحة الرئيسية للأخبار.

Выскажите своё мнение (по-английски).
Напишите нам
Подпишитесь на ленту новостей RSS.
Новые публикации
Новости главной страницы.

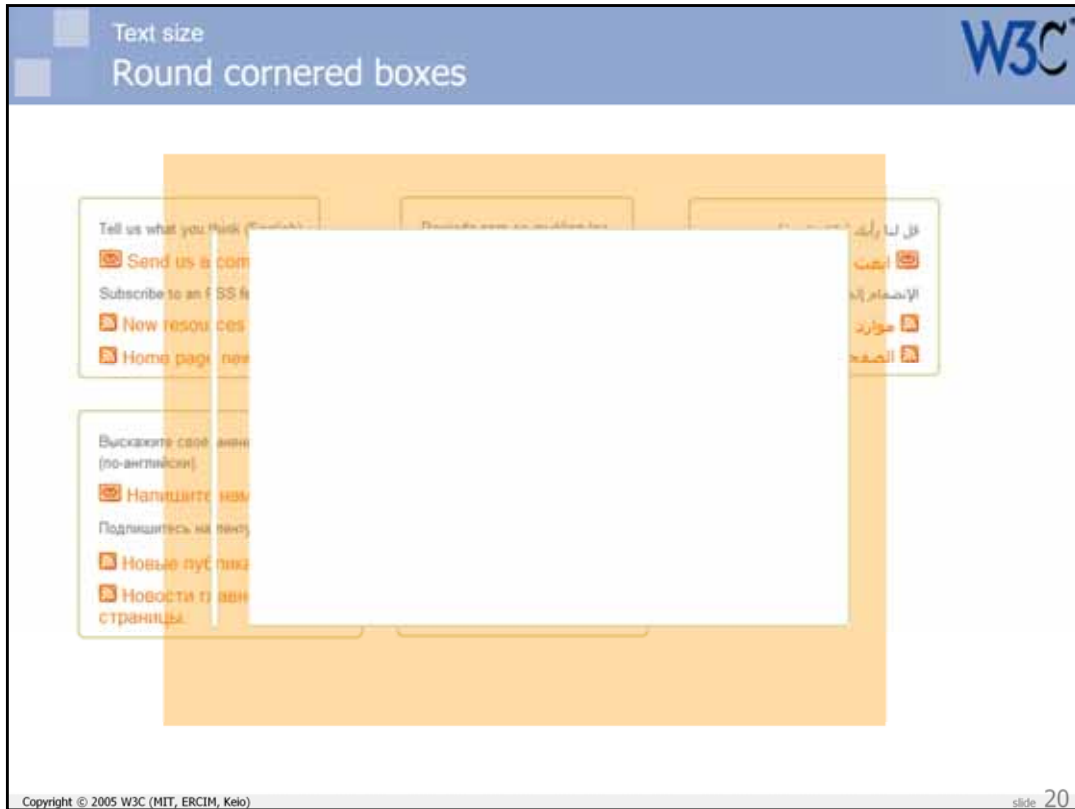
告诉我们您的想法 (用英语)。
将您的评论发送给我们
订阅 RSS 摘要
新的资源
主页新闻。

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 19

This will be much easier to do when CSS3's rounded corners properties are available. In the meantime, people are using background graphics.

This box can use techniques similar to those described earlier to allow the box to contract or expand (up to a certain point) horizontally as well as vertically, to fit the text.



This shows the graphics used.

Text size
Round cornered boxes

W3C

English:
Input processing features

German:
Eingabeverarbeitungsfunktionen

Eingabeverarbeitungsfunktionen

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 21

One potential issue here is that some languages have very long words, which may not fit within the specified max-width.

Text size
Round cornered boxes

W3C

English:
Input processing features

German:
Eingabeverarbeitungsfunktionen

German:
Eingabeverarbeitungs-
funktionen

Eingabe&00AD;verarbeitungs&00AD;funktionen

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 22


An ideal solution to this would be to get the translator to insert soft-hyphens at appropriate places in the word. The word would then wrap to fit the box.

Unfortunately, although this works in browsers like IE and Opera, the soft hyphen is not yet recognized in Mozilla browsers (although it should do in version 3).

Until it is available widely as a solution, the translator will probably need to break the text manually, and the designer should try not to make such boxes too narrow.

What this means to me

Some lessons from this section



- Think about how to cope with text that expands in translation
- Avoid placing labels to the side of form entry fields, icons, etc, if you can
- For text backgrounds use large images and 'sliding door' effects to allow coverage of text whatever the size
- Leave expansion space at the end of lines of tabs
- Consider how to deal with long compound nouns: don't make the column too narrow; warn translators
- Look out for and lobby the W3C CSS Working Group to publish rounded borders and multiple backgrounds parts of CSS3
- Lobby Mozilla to support the soft-hyphen character

<http://www.w3.org/International/questions/qa-resizing-backgrounds>

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

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

The slide features a blue header bar at the top. On the left side of the header, there are two small square icons, one above the other. To their right, the text "Navigating to localized pages" is displayed in white. On the right side of the header, the W3C logo is visible in white. The main body of the slide is white and contains the title "Navigating to localized pages" in a large, blue, sans-serif font, centered horizontally. At the bottom left of the slide, there is a small copyright notice: "Copyright © 2005 W3C (MIT, ERCIM, Keio)". At the bottom right, the text "slide 24" is present.

Navigating to localized pages
Language links

W3C

Google 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 pull-down 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	

Safe Search 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 25

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. 117 languages.

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.

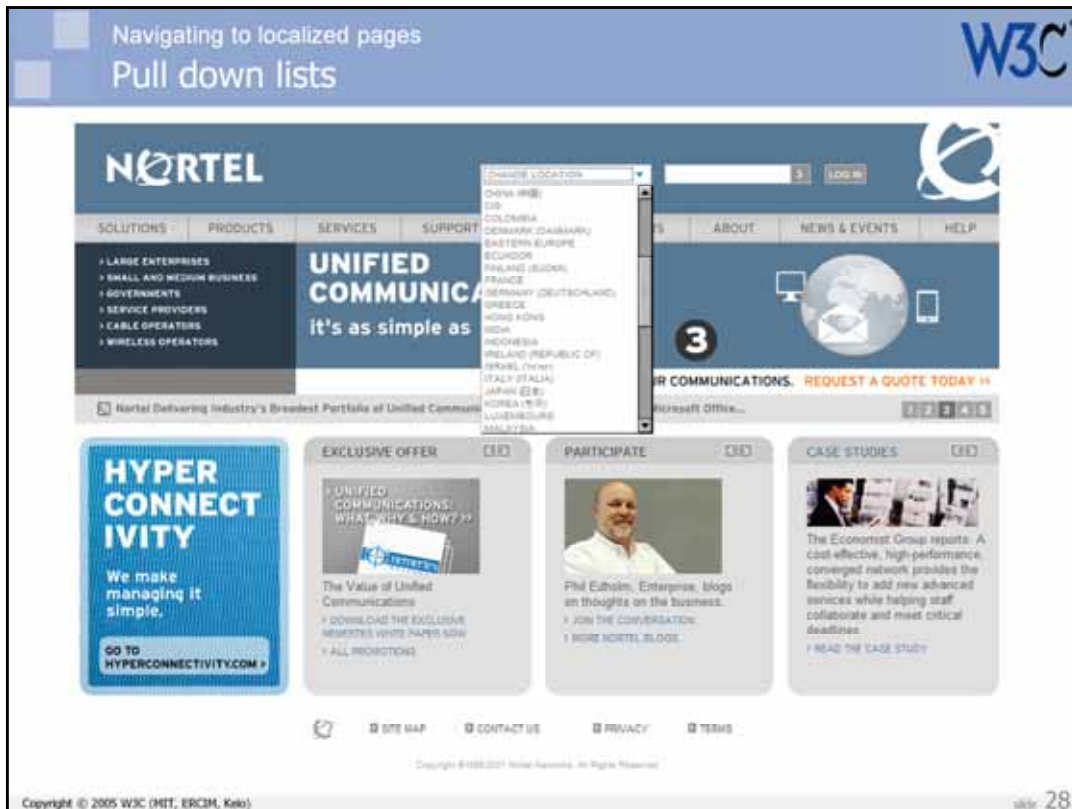


Note that the names of languages are all in Persian, and are sorted by Persian rules.

Which selection would get you back to English ??

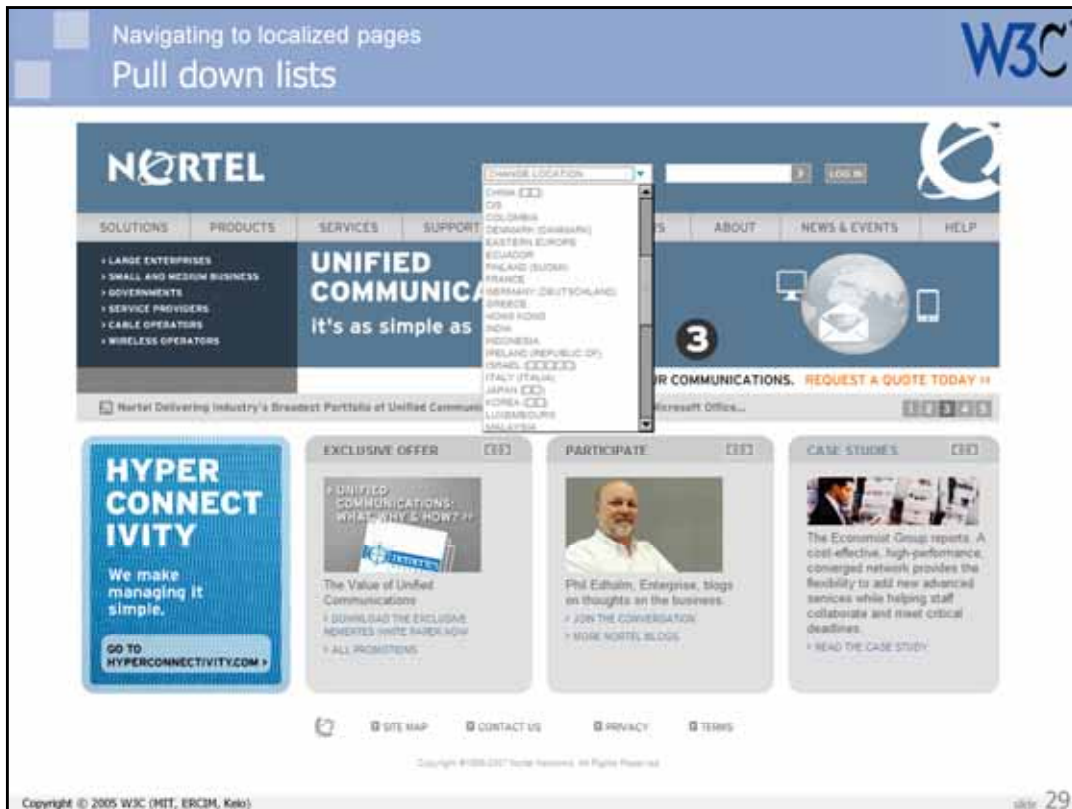
(Hint: if you want to explore like this, use a different tab or window for your explorations, and leave the original dialogue available in another for when you want to reset to your current language.)

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>.



One way of allowing users to access an alternative language or country site is to use pull-down lists (with the select element in HTML).

Here is an example. My personal preference would be to use the native translation of the country name first and put the English version in parentheses after. It can be useful, however, to have both as shown here.



A potential issue with this approach is that the user may not have fonts to cover all the characters, resulting in something like the boxes shown above.

You may feel that this is acceptable, since the English text alongside gives sufficient cues to the reader about what those boxes are about.

On the other hand, some people may feel that this is not acceptable.



An alternative approach is to route the user to a selection page like this. On such a page, which is in HTML, you can resolve the missing font issue by presenting the text in graphic form.

I would not normally recommend using text in graphics, but there shouldn't be a big issue here, given that this text doesn't need translating and is perhaps less likely to be searched for.

One thing to bear in mind, however, is that you should route a user to an alternative language and (often) country site that doesn't remove them from their current context, i.e. don't force them up to the home page if they entered the site from a link. If they go to a selection page like this, you should return them to the same page (if such exists) in the localized site.

Navigating to localized pages
Using alternative selection lists

The screenshot shows the Symantec website interface for a localized page. At the top, there is a navigation bar with the Symantec logo and a search bar. Below the navigation bar, there is a main content area with a large banner for 'Go Ahead. You've Got Norton.' and 'Introducing Norton 2008 Products'. To the left, there is a sidebar with various links and a pull-down menu that is open, showing a list of countries and languages. The pull-down menu includes options like 'Australia - English', 'Canada - English', 'France - French', 'Germany - German', 'India - English', 'Japan - Japanese', 'UK - English', 'USA - English', etc. The page footer contains copyright information: 'Copyright © 2005 W3C (MIT, ERCIM, Keio)'.

This example shows a hybrid approach, where what looks like a pull-down list has been implemented using javascript and HTML. Although they haven't done so in this case, this would allow you to use graphics in the pull-down.

W3C

 Navigating to localized pages
 Chatty selection sites

Country Sites

Either the webapp context your region JSP URI refers to doesn't exist or the calling web application does not have permission to access another web application context. If the web application exists in your application server, consult the Trouble Shooting Guide for details on how to configure your application server to support cross context forwarding.

American Samoa * English	Estonia * Estonian	Mexico * Spanish * English	Singapore * English
Argentina * Spanish	Fiji * English	Netherlands * Dutch	Slovak Republic * Slovak
Australia * English	Finland * Finnish	New Zealand * English	Slovenia * Slovenian
Austria * German	France * French	Nicaragua * Spanish	South Africa * English
Belgian * Dutch	French Polynesia * English	Naze * English	South Korea * Korean

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 32

If you take users to a selection page, don't be too chatty on that page. If they are going to it because they don't speak the current language, it won't be a lot of help.

Note that this particular introductory text is even difficult for English-speaking people to understand. Don't worry your non-English speaking people with such information.

■ Navigating to localized pages issues
W3C[®]

■ Identifying the list

The screenshot shows the Xerox website interface in Russian. At the top, there is a navigation bar with the Xerox logo and a search box. Below the navigation bar, there is a main content area with a large graphic of a document and a red banner with the text: "Добро пожаловать на новый виток Вашего успешного развития. Давайте начнем сочинять." (Welcome to a new chapter of your successful development. Let's start writing.) Below the banner, there are three columns of text: "Оборудование для дома и офиса", "Промышленная и коммерческая печать", and "Широкоформатные системы". On the left side, there is a "Новости" (News) section with several articles. On the right side, there is a "Гарантия сервис XEROX" (Xerox Service Guarantee) section. At the bottom, there is a footer with copyright information and a page number "33".

In this example, the link to the selection page (though it could also be the location of a pull-down list) is not easy to spot from the text. The designer has, however, used a graphic – ie. a non-verbal cue – to help users locate the appropriate place (albeit a little difficult to spot and recognize).

Navigating to localized pages
Using maps

W3C

The screenshot shows the Xerox website interface. At the top, there is a navigation bar with the Xerox logo and a search bar. Below the search bar, there are several menu items: Office Equipment, Production Equipment, Document Delivering, Supplies, Support & Drivers. A "Change Country" link is visible next to "United States".

The main content area is titled "Xerox Websites" and includes a link to "View a list of Worldwide Contacts". Below this is a world map with different regions highlighted in various colors. Underneath the map, there are two sections of localized links:

- North America:**
 - Canada/English
 - Canada/Français
 - United States of America
- Latin America and Caribbean:**
 - Argentina
 - Bermuda
 - Brazil
 - Cayman Islands
 - Chile
 - Colombia
 - Ecuador
 - Mexico
 - Honduras
 - Paraguay
 - Puerto Rico
 - Turkey & Tobago
 - Uruguay
 - Venezuela

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 34


One issue that is difficult to resolve is how to order the list of languages or countries. You could use the Unicode Collation Algorithm, sort by population size or GNP, sort by BCP 47 language tags, etc. to be neutral, although many users would possibly not understand the rationale for such ordering.

For **country-specific** links it may help to use maps of regions to narrow the search to a list that may be easier to scan.

Don't use flags for language-specific selections.

What this means to me

Some lessons from this section



- For links to alternative language or country sites, translate link text into the language of the content pointed to
- Make it easy to find the current page when changing language or site
- Consider whether to use graphics or text, given that fonts may not be available on the user's platform for all options
- If you use text, you'll need the page encoding to be UTF-8
- Remember that pull-down lists don't support graphics, so consider alternatives
- Find a way to draw attention to a list that is recognizable regardless of language, and put it in a visible location
- Consider maps as an alternative selection mechanism

<http://www.w3.org/International/questions/qa-navigation-select>

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

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

Composite messages

W3C

Composite messages

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

This slide features a blue header bar with the text "Composite messages" on the left and the W3C logo on the right. The main content area is white and contains the text "Composite messages" centered. The footer contains copyright information and the slide number "36".

Composite messages
Syntax differences

W3C

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。

يكلّمونها من بيتي الجديد

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

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

Composite messages
Agreement & Concept mappings

W3C

On

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

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

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.

Composite messages

Agreement of Russian plurals

W3C

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

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

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

Composite messages
Handling variables

W3C

W3C **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/20060909 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.

Jump To: Results

This page is not Valid XHTML 1.0 Transitional!

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

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

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

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 40

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.

Composite messages
Handling variables

W3C

W3C **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_115929539132.html

Modified: jundfinazf

Server: Microsoft Windows [Windows NT 5.1 x64]
C:\Program Files\Microsoft Windows\

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 41

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

Composite messages
Handling variables

W3C

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: webdevs:/localhost_1159295639132.html

Modified: (undefined)

Screen: Microsoft Windows [Version 5.0.5260.1713] (x86) : Internet Explorer 6.0.2600.5512

Jump To: Results

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 42

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."

Composite messages
Handling variables

W3C

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: webdevs:/localhost_1159295639132.html

Modified: (undefined)

Screen: Microsoft Internet Explorer 6.0.2.6002

Jump To: Results

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 43

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

Composite messages
Handling variables

W3C

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: webdevs_localhost_1159293639132.html

Modified: (undefined)

Screen: Microsoft Windows [Version 5.0.5260.1713] (c) 2006 Microsoft Corporation. All rights reserved. C:\Program Files\Internet Explorer\iexplore.exe

Jump To: Results

Die Datei 268 enthält 0 Gültigkeitsfehler. X

```
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 44

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.

Composite messages
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: webdevs/localhost_1159293639132.html

Modified: (undefined)

Screen: Microsoft Windows [Version 5.0.2600.5512] Copyright (c) 2006 Microsoft Corporation. All rights reserved. C:\Program Files\Internet Explorer\iexplore.exe

Jump To: Results

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 45

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.

Composite messages
Handling variables

W3C

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: webdevs_localhost_1159293639132.html
Modified: (undefined)
Server: Microsoft IIS/5.0.5260.5500 (localhost/localhost)

Jump To: Results

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 46

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.

Composite messages
Handling variables

W3C

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: webdev..._localhost_1159295639132.html
Modified: jun05/04
Server: Apache/2.0.46 (Ubuntu Linux) PHP/5.0.3

Jump To: Results

1 ошибка
2 ошибки
5 ошибок

При проверке файла myFirst.html была обнаружена 268 ошибка.

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

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

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

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.

Composite messages
Handling variables

W3C

W3C **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: webdev/_localhost_/1159295639132.html

Modified: jun06/06

Server: Apache/2.0.46 (Ubuntu) PHP/5.1.4-1ubuntu4.1

Jump To: Results

1 ошибка
2 ошибки
5 ошибок

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."
)
```

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

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.

Composite messages
Sentence-like composite messages

W3C

The < > is on.

printer stacker

 stapler options

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

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).

Composite messages
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 50

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/> .

Composite messages
Text string re-use

W3C

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

Printer: On

Stacker: On

Stapler options: On

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

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.

Composite messages
Text string re-use

W3C

```
<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 52

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/>.

What this means to me

Some lessons from this section

W3C

- 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

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 53

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


Concluding remarks

W3C

Concluding remarks

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

This slide features a blue header bar with the text "Concluding remarks" on the left and the W3C logo on the right. The main content area is white and contains the text "Concluding remarks" centered. The footer contains copyright information on the left and "slide 54" on the right.




Quick tips

- **Encoding.** Use Unicode wherever possible for content, databases, etc. Always declare the encoding of content.
- **Escapes.** Use characters rather than escapes (e.g. á á or á) whenever you can.
- **Language.** Declare the text-processing language of documents and indicate any internal language changes.
- **Presentation vs. content.** Use style sheets for presentational information. Restrict markup to semantics.
- **Images, animations & examples.** Check for translatability and inappropriate cultural bias.

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

This and the following slide contain the text on our Quick Tips cards. You can get more information on these topics at <http://www.w3.org/International/quicktips/>.




Quick tips

- **Forms.** Use an appropriate encoding on both form and server. Support local formats of names/addresses, times/dates, etc.
- **Text authoring.** Use simple, concise text. Use care when composing sentences from multiple strings.
- **Navigation.** On each page include clearly visible navigation to localized pages or sites, using the target language.
- **Right-to-left text.** For XHTML, add `dir="rtl"` to the html tag. Only re-use it to change directionality.
- Check your work. Validate! Use techniques, tutorials, and articles at <http://www.w3.org/International/>

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

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 56




In conclusion

- We haven't been talking about things that the localizers do, we're talking about the way content is created!
- Good internationalization practice is often just good content design.
- This is a part of Web Accessibility.
- Take a look at the material on the I18n Activity site.

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

The presentation is not designed to give you a thorough overview of potential internationalization and localization issues. 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 designers out of their comfort zone, and help them 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.



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

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 58