



Practical Tips for Designing International Web Pages

Richard Ishida
W3C Internationalization Activity Lead

The slide features a world map in shades of blue and grey. Three teal-colored text boxes are overlaid on the map. The top-left box contains a small portrait of Sir Tim Berners-Lee and text about the W3C's creation. The top-right box contains the organization's mission statement. The bottom-center box contains key terms. The W3C logo is in the top right corner, and copyright information is in the bottom left.

About the W3C

W3C[®]

1994: World Wide Web Consortium created and still led by: **Sir Tim Berners-Lee**, inventor of the WWW.

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

Keywords: **consensus** and **vendor neutrality**

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

About the W3C W3C[®]

Interaction
Graphics
HTML
Internationalization
Math
Rich Web Client
Style
Synchronized Multimedia
Video in the Web
XForms

Web Accessibility
International Program Office
Technical Activity

Technology & Society
eGovernment
Patent Policy
Privacy
Security
Semantic Web
Web Services

Ubiquitous Web
Mobile Web
Multimodal Interaction
Ubiquitous Web Apps
Voice Browser
XML

4 Domains
22 Activities
50 Working Grp
13 Interest Grp
5 Coordination Grp

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

Outline

- Text size
- Navigating to localized pages
- Separating content & presentation
- Composite messages

International Web Sites

Text size

W3C[®]

Text size

Copyright © 2005 W3C (MIT, ERCIM, Keio)


slide 5

International Web Sites

Text size
Text size in translation

W3C[®]

DochuLa Pass



Around 3,100 m high, in Bhutan.

Anyone can see this photo ([edit](#))
uploaded on Jan 28, 2007 | [Map](#) | [Delete](#)
1,763 views 5 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 6

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

Text size
Text size in graphics


W3C

Seals and sealions at Pier 39



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

Seals and sealions at Pier 39



ADICIONAR OBSERVAR EN UN GRUPO ENVIAR PARA ADICIONAR A LA BARRA DE BLOQUEAR TODOS TAMAÑOS IMPRIMIR GIRAR BORRAR X


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

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.

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 8

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[®]

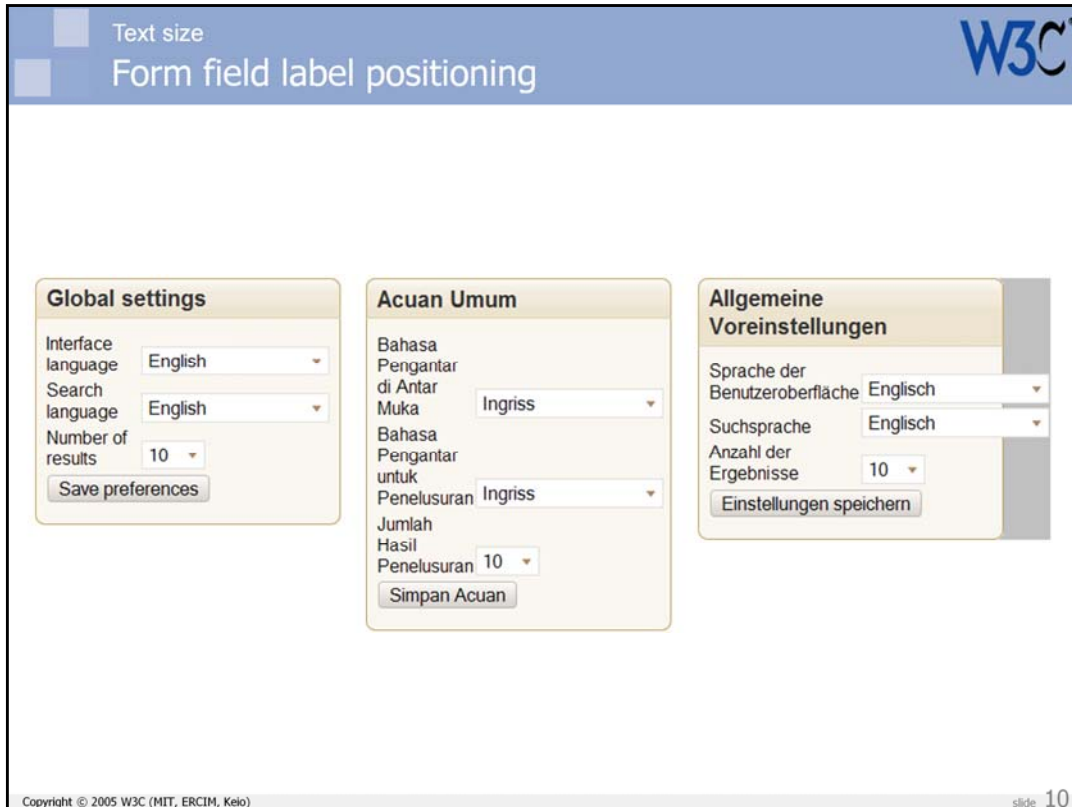
Global settings

Interface language	English ▾
Search language	English ▾
Number of results	10 ▾
<input type="button" value="Save preferences"/>	

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

Here is another potential issue related to the length of text in translation. 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 continue to assume a situation where text appears in a fixed width box. We have used a table to apply form labels to the left side of the form entry field to which they apply. The table framework should allow text to expand vertically without destroying the relationship with the input controls. Our initial source text is in English.



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.

(Note that these are screen snaps – ie. they are not theoretical problems.)

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
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 11

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 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.

Text size

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 12

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

(This example is in Spanish, and 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 13

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.


Text size

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;  
}
```

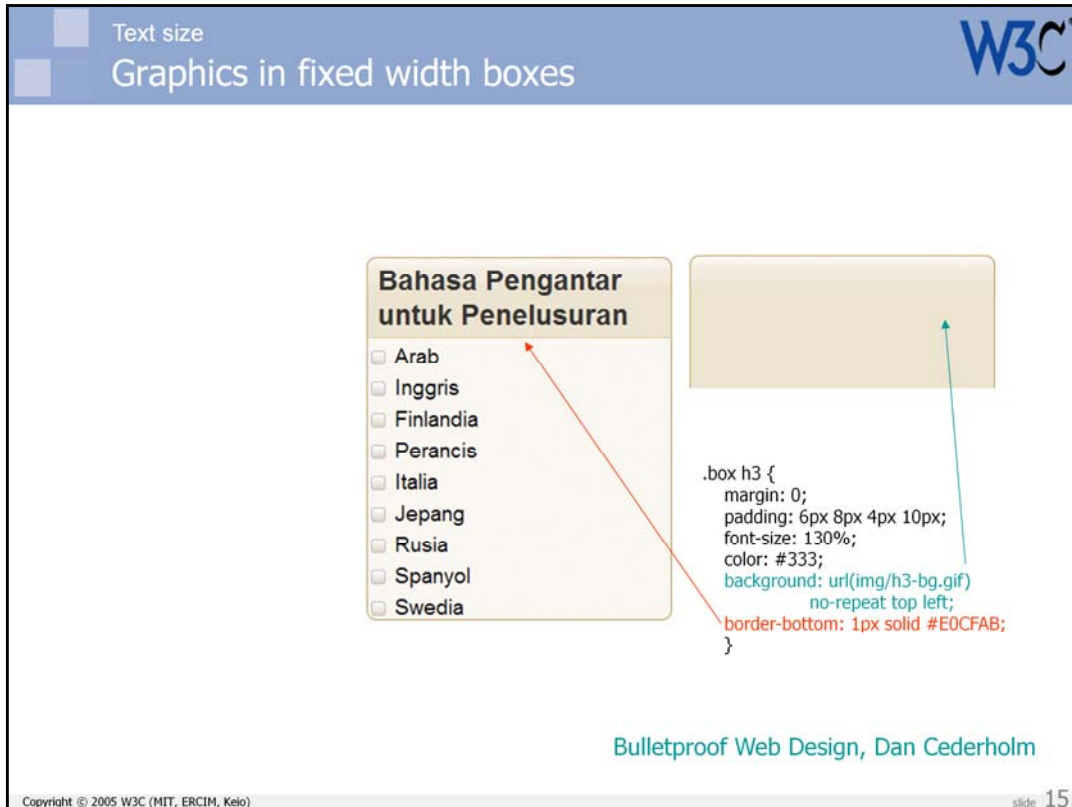
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 14

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

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;  
}
```

Bulletproof Web Design, Dan Cederholm

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 15

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.

International Web Sites

Text size
Tabs

W3C

首頁 你的足跡 管理 自己人 小組

Início Minhas coisas Organizar Amigos Grupos

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

When creating tabs, don't draw the background as a single graphic.

The image shows a screenshot of a web page with a navigation menu. At the top left, there are controls for 'Text size' and 'Tabs'. The W3C logo is in the top right. The navigation menu consists of two rows of tabs. The top row has five tabs with Chinese text: '首頁', '你的足跡', '管理', '自己人', and '小組'. The bottom row has five tabs with English text: 'Home', 'Tu', 'Organizza', 'Contatti', and 'Gruppi'. The 'Tu' tab is currently selected. Below the tabs is a large, empty rectangular area. At the bottom right, there is a link to 'Sliding Doors of CSS, Doug Bowman' with the URL 'http://alistapart.com/articles/slidingdoors/'. The bottom left corner contains the copyright notice 'Copyright © 2005 W3C (MIT, ERCIM, Keio)' and the bottom right corner shows 'slide 17'.

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.

Text size
Tabs



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

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 "undo" two-character year fields that were built on the assumption of "190x").

So ideally, 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 [qa-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 [Member](#) privacy statements.

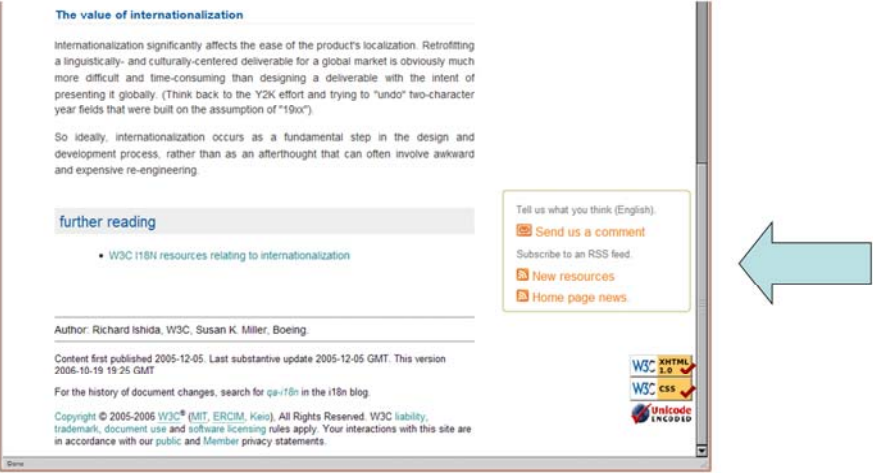
Tell us what you think (English).

- Send us a comment
- Subscribe to an RSS feed.
- New resources
- Home page news

W3C XHTML 1.0

W3C CSS

Unicode



slide 19

Copyright © 2005 W3C (MIT, ERCIM, Keio)

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

Text size

Round cornered boxes

W3C[®]

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 (بالإنجليزية).
✉ ابعث تعليق.
الإضمام 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 20

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.

International Web Sites

Text size

Round cornered boxes

W3C[®]

The screenshot displays a web page with content in three languages: English, Arabic, and Russian. The English section includes links for 'Send us a comment', 'Subscribe to an RSS feed', 'New resources', and 'Home page news'. The Arabic section includes 'أرسل لنا رأيك', 'أرسل لنا تعليقاً', 'الانضمام إلى الإصدار الإلكتروني', 'مصادر', and 'الصفحة الرئيسية'. The Russian section includes 'Выскажите свое мнение (по-английски)', 'Напишите нам', 'Подпишитесь на нашу рассылку', 'Новые публикации', and 'Новости главной страницы'. A large white rectangular box is overlaid on the page, highlighting the 'Send us a comment' link in the English section.

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

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 22

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:
Eingabeverarbeitungs-funktionen

German:
Eingabeverarbeitungs-
funktionen

Eingabe&00AD;verarbeitungs&00AD;funktionen


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

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.

(Although this works in the latest versions of all major browsers, the soft hyphen was only implemented in Mozilla browsers from Firefox version 3.)

What this means to me

Some lessons from this section

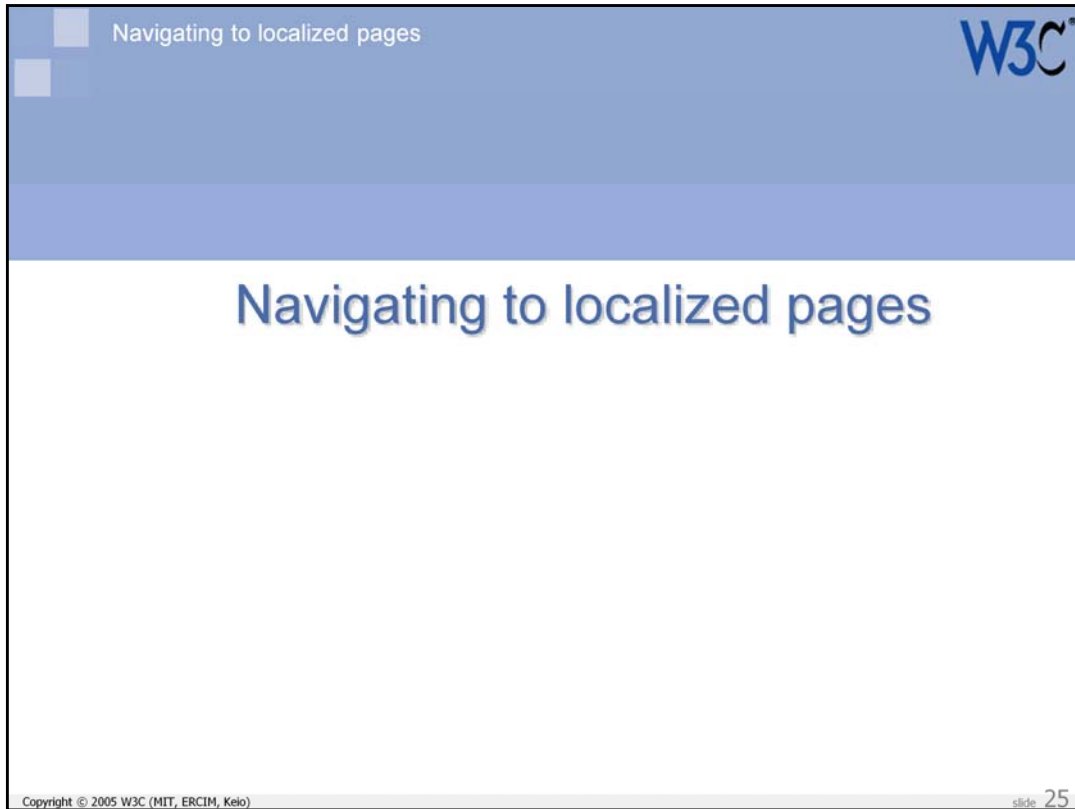


- Think about how to cope with text that expands in translation
- Avoid placing labels to the side of form entry fields, icons, etc, if you can
- For text backgrounds use large images and 'sliding door' effects to allow coverage of text whatever the size
- Leave expansion space at the end of lines of 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 browser developers to support internationalization features such as the soft-hyphen character


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

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

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



Navigating to localized pages



Navigating to localized pages

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 25

Navigating to localized pages
Language links

W3C[®]

The screenshot shows the Google Preferences page. At the top, it says 'Google Preferences' with links for 'Preferences Help' and 'About Google'. Below this is a 'Save your preferences when finished and return to search.' button and a 'Save Preferences' button. The main section is 'Global Preferences (changes apply to all Google services)'. Under 'Interface Language', there is a dropdown menu set to 'English'. Under 'Search Language', there are two radio buttons: 'Search for pages written in any language (Recommended)' (which is selected) and 'Prefer pages written in these language(s)'. Below these are checkboxes for various languages including Afrikaans, Arabic, Armenian, Belarusian, Bulgarian, Catalan, Chinese (Simplified), Chinese (Traditional), Croatian, Czech, Danish, Dutch, English, Esperanto, Estonian, Filipino, Finnish, French, German, Greek, Hebrew, Hungarian, Icelandic, Indonesian, Italian, Japanese, Korean, Latvian, Lithuanian, Norwegian, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swahili, Swedish, Thai, Turkish, Ukrainian, and Vietnamese. Under 'SafeSearch Filtering', there are three radio buttons: 'Use strict filtering (Filter both explicit text and explicit images)', 'Use moderate filtering (Filter explicit images only - default behavior)' (which is selected), and 'Do not filter my search results.'. Under 'Number of Results', it says 'Google's default (10 results) provides the fastest results.' and 'Display 10 results per page.'.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 26

Note, in passing, an issue related to the Google text I used in an earlier 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.

Navigating to localized pages
Language links

W3C[®]

The screenshot shows the Google Preferences page. At the top, there is a header with the text 'Navigating to localized pages' and 'Language links', and the W3C logo. Below this is the Google Preferences interface. The 'Global Preferences' section is active, showing 'Interface Language' set to 'English'. A dropdown menu is open, displaying a list of languages including English, Persian, and many others. The 'Search Language' section has two radio buttons: 'Search for pages written in any language' (selected) and 'Prefer pages written in these languages'. Below this is a grid of checkboxes for various languages. The 'SafeSearch Filtering' section has three radio buttons: 'Use strict filtering', 'Use moderate filtering' (selected), and 'Do not filter my search results'. The 'Number of Results' section shows 'Display 10 results per page'. At the bottom left, there is a copyright notice: 'Copyright © 2005 W3C (MIT, ERCIM, Keio)'. At the bottom right, there is a slide number: 'slide 27'.

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

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.

Navigating to localized pages
Language links

W3C

The screenshot shows the Persian version of the Google homepage. At the top, there is a search bar and the Google logo. Below the search bar, there is a dropdown menu for language selection. The menu is currently set to Persian. A list of language options is displayed, including English, Spanish, French, German, Italian, Japanese, Korean, Chinese, Hindi, and many others. The interface is in Persian script.

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

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.



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>.

W3C[®]

Navigating to localized pages
In-page links

The screenshot shows a W3C page titled "Character encodings" under the "Internationalization" section. A language menu is overlaid on the right side of the page, listing languages: Български, Español, Français, Polski, Português-BR, Русский, Svenska, ไทย, and Українська. A light blue arrow points to this menu. The page content includes sections for "the document character set" and "declaring encodings".

Internationalization

Home Resources Techniques

Character encodings

about

on this page: document character set · declaring encodings · ensuring de further reading

Intended audience: anyone wanting a quick summary of key information relat and XML. For more information follow the links in the text or see Further reading

the document character set

The document character set for XML and HTML 4.0 is Unicode (aka ISO browsers and XML processors should behave as if they used Unicode in documents have to be transmitted in Unicode. As long as client and ser can use any encoding that can be converted to Unicode. Read more abou

declaring encodings

It is very important that the character encoding of any XML or XHTML document is clearly labeled, so that clients can easily map these encodings to Unicode. This can be done in the following ways:

- Send the 'charset' parameter in the Content-Type header of HTTP. Example:

```
Content-Type: text/html; charset=utf-8
```

To do this you will need to have access to server settings or serve your document via scripting (see Setting the HTTP charset parameter for more information).
- For XML (including XHTML), use the encoding pseudo-attribute in the XML declaration at the start of a document or the text declaration at the start of an entity. Example:

```
<?xml version="1.0" encoding="utf-8" ?>
```

related links

Resources

118n site search:

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 30

Ideally you would display individual links to translated pages on the page itself.

These should be in a prominent place so that they can be easily found.

I recommend that you always provide these links, regardless of whether you try to automatically guide the user to the right page using content negotiation, ip lookup, etc.

International Web Sites

Navigating to localized pages
Pull down lists

The screenshot shows the Nortel website interface. At the top right, there is a 'W3C' logo. Below it, a blue header contains the text 'Navigating to localized pages' and 'Pull down lists'. The main content area features the Nortel logo and a navigation menu with categories: SOLUTIONS, PRODUCTS, SERVICES, and SUPPORT. A 'CHANGE LOCATION' pull-down menu is open, displaying a list of countries and regions: CHINA (中国), CIS, COLOMBIA, DENMARK (DANMARK), EASTERN EUROPE, ECUADOR, FINLAND (SUOMI), FRANCE, GERMANY (DEUTSCHLAND), GREECE, HONG KONG, INDIA, INDONESIA, IRELAND (REPUBLIC OF), ISRAEL (ישראל), ITALY (ITALIA), JAPAN (日本), KOREA (한국), LUXEMBOURG, and MALAYSIA. Below the menu, there are several promotional banners: 'HYPER CONNECTIVITY', 'EXCLUSIVE OFFER', 'PARTICIPATE', and 'CASE STUDIES'. The footer contains links for 'SITE MAP', 'CONTACT US', 'PRIVACY', and 'TERMS', along with copyright information: 'Copyright © 2005 W3C (MIT, ERCIM, Keio)'. A slide number '31' is visible in the bottom right corner.

If you have a large number of options, 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 (though inconsistently). My preference would also be to put the non-English first, and add the English in parentheses after (unlike this example).

International Web Sites

Navigating to localized pages
Pull down lists

Copyright © 2005 W3C (MIT, ERCIM, Keio)

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.

Navigating to localized pages
Using alternative selection lists

Copyright © 2005 W3C (MIT, ERCIM, Keio)

This example shows an approach that could be used (although hasn't here) to provide what looks like a pull-down list using javascript and HTML. This produces a hybrid approach that would allow you to use graphics in the pull-down, so that all native language text could be seen as intended.

International Web Sites

W3C

Navigating to localized pages
Selection pages

Search | A-Z Index | Feedback | **Languages**

WELCOME

The Church of Jesus Christ of Latter-day Saints

Home | **Languages**

Choose a Language

- Bahasa Indonesia
- Česky
- Dansk
- Deutsch
- Español
- Eesti keeles
- Français
- Gagana Samoa
- Hiligaynon
- Hmoob
- Hrvatski
- Ilokano
- Italiano
- Kiribati
- Kreyòl Ayisyen
- Lea Fakatonga
- Magyar
- Nederlands
- Norsk
- Polski
- Português
- Reo Tahiti
- Română
- Slovenčina
- Slovensčina
- Suomi
- Svenska
- Tagalog
- Tiếng Việt
- Vosa vakaviti
- Ελληνικά
- Български
- Монгол
- Русский
- Українська
- ภาษาไทย
- ភាសាខ្មែរ
- ភាសាថៃ
- 國語
- 廣東話
- 日本語
- 한국말

© 2005 Intellectual Reserve, Inc. All rights reserved. [Rights and use information](#) [Privacy policy](#)

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

An alternative approach is to route the user to a selection page like this. On such a page, which is in HTML, you can also 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, ie. 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.


W3C[®]

Navigating to localized pages issues
Identifying the list

The screenshot shows the Xerox website interface in Russian. At the top, there is a search bar with the text "Выберите регион" (Select region) and a "Искать" (Search) button. Below the search bar is a navigation menu with links: "О XEROX", "Контакты", "Правос-Центр", "Для партнеров", "Где купить", and "Сервисные центры". The main content area features a large 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 are three columns of product categories: "Оборудование для дома и офиса" (Home and office equipment), "Производительная и коммерческая печать" (Productive and commercial printing), and "Широкоформатные системы" (Wide format systems). There are also sections for "Новости" (News) and "Публикации" (Publications). At the bottom, there is a footer with copyright information: "Copyright © 2005 W3C (MIT, ERCIM, Keio)" and "slide 35".

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.

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
Belgium ▪ Dutch	French Polynesia ▪ English	Niue ▪ English	South Korea ▪ Korean

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

If you do 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.

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, a search bar, and a 'Change Country' link. Below the navigation bar, there is a section titled 'Xerox Websites' with a link to 'Worldwide Contacts'. A world map is displayed, with different regions highlighted in various colors. Below the map, there are two main categories of localized links: 'North America' and 'Latin America and Caribbean'. Each category has a small map icon and a list of links for specific countries or regions.

United States [Change Country](#)

XEROX

[Where to Buy](#) | [Contact Us](#) | [Log In/My Xerox](#) | [Account Management](#) | [Shopping Cart](#)

[Office Equipment](#) | [Production Equipment](#) | [Document Outsourcing](#) | [Supplies](#) | [Support & Drivers](#)

[Contact Xerox](#)

Xerox Websites

View a list of [Worldwide Contacts](#).

North America
[Canada/English](#) | [Canada/Français](#) | [United States of America](#)

Latin America and Caribbean
[Argentina](#) | [Chile](#) | [Nicaragua](#) | [Uruguay](#)
[Bermuda](#) | [Colombia](#) | [Paraguay](#) | [Venezuela](#)
[Brasil](#) | [Ecuador](#) | [Peru](#)
[Cayman Islands](#) | [México](#) | [Trinidad & Tobago](#)

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


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 for regional sites. Don't use flags for languages.

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

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

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

Separating content & presentation

W3C[®]

Separating content & presentation

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

The slide features a blue header bar with the text 'Separating content & presentation' on the left and the W3C logo on the right. The main body of the slide is white with the same text 'Separating content & presentation' centered in a large blue font. The footer contains copyright information on the left and 'slide 39' on the right.

W3C[®]
Separating content & presentation
Practical implications

Յունիկոդը փոխում է այդ ամենը

Յունիկոդը յուրաքանչյուր նիշի ուրույն կող է վերագրում՝ ծրագրահենքից, ծրագրից և լեզվից անկախ: Յունիկոդի ստանդարտն ընդունվել է ոլորտի այնպիսի առաջատարների կողմից, ինչպիսիք են Apple, HP, IBM, JustSystems, Microsoft, Oracle, SAP, Sun, Sybase, Unisys և [բազմաթիվ այլ](#) ընկերություններ:...

```

<p><font size="3"><i>Յունիկոդը փոխում է այդ
ամենը</b></i></font></p>
```

```

<p>Յունիկոդը յուրաքանչյուր նիշի ուրույն կող է վերագրում՝
ծրագրահենքից, ծրագրից և լեզվից անկախ: Յունիկոդի ստանդարտն
ընդունվել է ոլորտի այնպիսի առաջատարների կողմից, ինչպիսիք են Apple,
HP, IBM, JustSystems, Microsoft, Oracle, SAP, Sun, Sybase, Unisys և <a
href="memblogo.html" style="font-style: italic;">բազմաթիվ այլ</a>
ընկերություններ:...
```

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

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

The basic recommendation is to avoid presentationally-oriented markup, such as or <i> tags, and inline style attributes. A key reason for this is that when you want to change the styling, it is too expensive to have to trawl through the HTML changing markup in every page.


Յունիկոդը փոխում է այդ ամենը


Յունիկոդը յուրաքանչյուր նիշի ուրույն կոդ է վերագրում՝ ծրագրահենքից, ծրագրից և լեզվից անկախ: Յունիկոդի ստանդարտն ընդունվել է ոլորտի այնպիսի առաջատարների կողմից, ինչպիսիք են Apple, HP, IBM, JustSystems, Microsoft, Oracle, SAP, Sun, Sybase, Unisys և [բազմաթիվ այլ](#) ընկերություններ:...

`<h3>Յունիկոդը փոխում է այդ ամենը</h3>`

`<p>Յունիկոդը յուրաքանչյուր նիշի ուրույն կոդ է վերագրում՝ ծրագրահենքից, ծրագրից և լեզվից անկախ: Յունիկոդի ստանդարտն ընդունվել է ոլորտի այնպիսի առաջատարների կողմից, ինչպիսիք են Apple, HP, IBM, JustSystems, Microsoft, Oracle, SAP, Sun, Sybase, Unisys և`

A better way is to use semantic markup (ie. meaning related markup) and relegate all styling considerations to a separate style sheet.

Separating content & presentation




```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
<title>About the W3C I18n Activity</title>
<style type="text/css" src="mystyling.css" />
</head>
<body>
<h1>I18n Activity, W3C</h1>
<div class="international-text" xml:lang="zh-Hans">国际化活动
<div class="description">
<p>The W3C Internationalization A
and coordinating any techniqu
activities within the W3C and t
that allow and make it easy to
with different languages, scrip
<p>The Activity comprises three W
(Guidelines, Education & Outr
Tag Set). There is also an inte
</div>
</body>
</html>
                    
```

```

body {
background: white;
color: black;
font-family: serif;
font-size: 1em;
}
h1 {
font-size: 240%;
}
div.international-text {
font-family: MingLiu, sans-serif;
font-size: 240%;
}
p {
margin-top: 1em;
}
                    
```

Content (XHTML)

Presentation (CSS)

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

In the HTML to the right of this slide there is no presentational information in the HTML – which is as it should be. Browsers have default styling built in, which would produce something like the layout on the left in Firefox.

To the right is some CSS code in a style sheet that can be used to apply styling to the HTML.

W3C®

Separating content & presentation

```

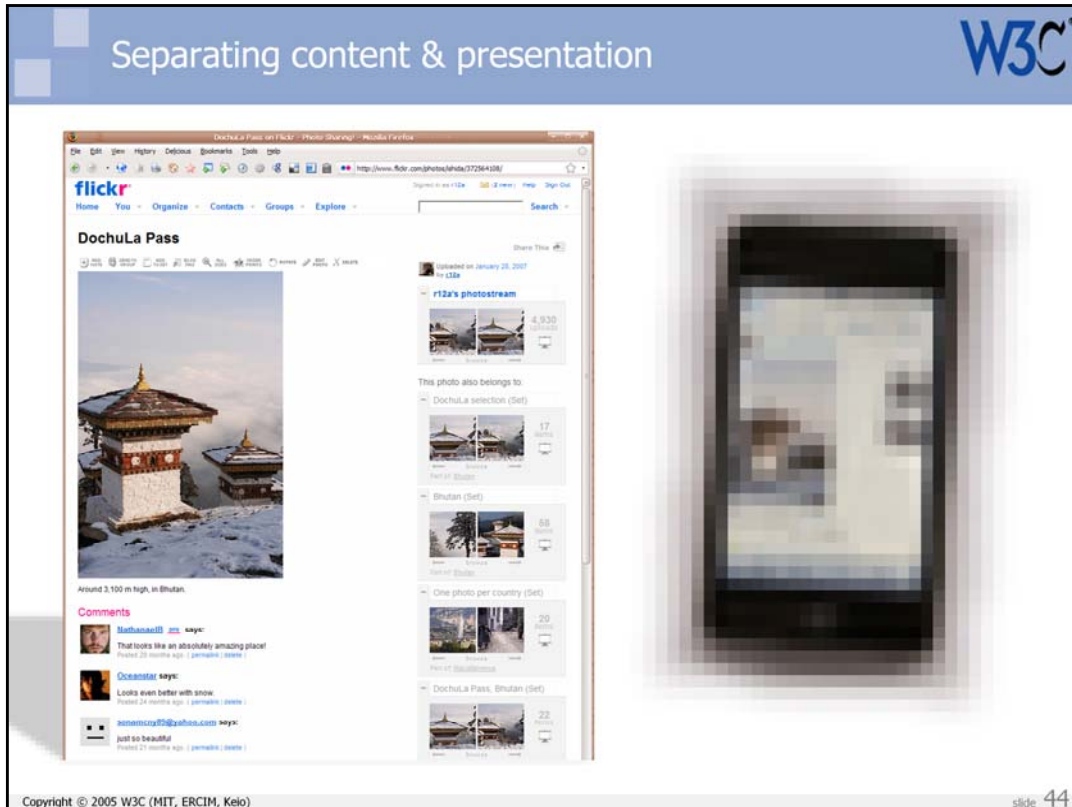
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
<head>
<meta content="text/html; charset=utf-8" />
<title>118n Activity</title>
<link href="mystyling.css" />
</head>
<body>
<div style="background-color: #e0e0e0; padding: 10px; width: 50%; float: left; margin-right: 10px;">
<h3 style="text-align: center; margin: 0;">国际化活动万维网联盟
<p style="margin: 0;">The W3C Internationalization Activity has the goal of proposing and coordinating any techniques, conventions, guidelines and activities within the W3C and together with other organizations that allow and make it easy to use W3C technology worldwide, with different languages, scripts, and cultures.
<p style="margin: 0;">The Activity comprises three Working Groups: Core, GEO (Guidelines, Education & Outreach), and ITS (Internationalization Tag Set). There is also an Internationalization Interest Group.
</div>
<div style="background-color: #e0e0e0; padding: 10px; width: 40%; float: right; margin-top: 10px;">
<h3 style="text-align: center; margin: 0;">118n Activity, W3C
<p style="margin: 0;">The W3C Internationalization Activity has the goal of proposing and coordinating any techniques, conventions, guidelines and activities within the W3C and together with other organizations that allow and make it easy to use W3C technology worldwide, with different languages, scripts, and cultures.
<p style="margin: 0;">The Activity comprises three Working Groups: Core, GEO (Guidelines, Education & Outreach), and ITS (Internationalization Tag Set). There is also an Internationalization Interest Group.
</div>
</body>
</html>
    
```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 43

The CSS style sheet is a single file that can be used by many markup files. Any changes to styling are therefore made in a single location.

Simple changes to the style sheet can be made to produce all the of layouts shown here, without any need to touch the HTML code at all.

Separating content & presentation W3C[®]



The image shows a desktop browser window displaying a Flickr page for 'DochuLa Pass'. The page features a large photo of a traditional Bhutanese building, a list of comments, and a sidebar with related photo sets. To the right of the browser window is a blurred image of a mobile phone screen, illustrating the concept of content separation and adaptation for different devices.


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

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

Pages tailored for mobile devices tend to use different styling for the same content.

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

Separating content & presentation W3C[®]



The image shows a desktop browser window displaying the Twitter homepage. The browser's address bar shows 'http://twitter.com/'. The page layout includes a search bar at the top, a 'What are you doing?' input field with a '140' character count, and a list of tweets. The tweets include news items from BBC News and other users. On the right side, there is a sidebar with user information for 'r12a' and a 'Following' list. To the right of the browser window is a blurred image of a mobile phone screen, illustrating how the same content is presented differently on a smaller device.

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

When it comes to aggregated data, different clients may use different styling for the same content.

Separating content & presentation
Style issues for Japanese

W3C

- emphasis techniques may be different because of problems to support bold and italics in small characters on-screen

これは日本語です。
これは日本語です。


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

Localization may also involve a need to restyle parts of the content.

Here are some ways in which typographic differences may appear between language versions of the same content, based on differences between English and Japanese. 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



- ◆ emphasis techniques may be different because of problems to support bold and italics in small characters on-screen
- ◆ 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 ที่เต็มไป ด้วยศักยภาพ

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

- emphasis techniques may be different because of problems to support bold and italics in small characters on-screen
- 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

经验分
(万维)

经验分
(万维)

日在德国
Mainz 市举行

日在德国 Mai
nz市举行的第

- emphasis techniques may be different because of problems to support bold and italics in small characters on-screen
- 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
- conventions differ: no upper- vs. lower-case distinction, no distinction between proportional and mono-spaced fonts

Press the START button.

Separating content & presentation
Don't forget your scripting

W3C

I18N Tests: Text-transform

This test examines how a user agent handles the CSS text-transform property across the full range of characters in Unicode's Basic Multilingual Plane that have upper/lower case equivalents.

Note: The CSS2 specification only requires that text-transform work for 'the Latin1 repertoire'. It is not clear what characters the term 'Latin1' covers. This test assumes that it means the Basic Latin and Latin-1 Supplement of the Unicode Standard (equivalent to ISO-8859-1). There is one character in this set, `y`, that only occurs in its lower-case form. Many browsers, however, support far more case conversions than required in the specification. In the following, the source contains pairs of characters, an uppercase character followed by a lowercase character, eg |A|a|B|b|C|c|. If the transform works correctly both characters should look the same. Where differences occur, view the source to see which characters are involved. (Tip: copy and paste the sequence into a plain text editor, or into Richard Ishida's [conversion page](#) or his [IPv6 view page](#).) On some user agents you will only see glyphs supported by the current font. To help with this situation, you can apply different fonts by typing in a different font name in the following box:

Latin1
 The css says: text-transform: uppercase.
 The right hand character should match the left hand one.
 A A | B B | C C | D D | E E | F F | G G | H H | I I | J J | K K | L L | M M | N N | O O | P P | Q Q | R R | S S | T T | U U | V V | W W | X X | Y Y | Z Z | A A | A A | A A | A A | A A | A A | A A | A A | E E | C C | E E | E E | E E | E E | I I | I I | I I | I I | D D | R R | O O | O O | O O | O O | O O | U U | U U | U U | U U | Y Y | P P | Y Y

The left hand character should match the right hand one.
 a a | b b | c c | d d | e e | f f | g g | h h | i i | j j | k k | l l | m m | n n | o o | p p | q q | r r | s s | t t | u u | v v | w w | x x | y y | z z | A A | A A | A A | A A | A A | A A | A A | A A | E E | C C | E E | E E | E E | I I | I I | I I | I I | D D | R R | O O | O O | O O | O O | O O | U U | U U | U U | U U | Y Y | P P | Y Y

General case: Latin
 The css says: text-transform: uppercase.
 The right hand character should match the left hand one.
 Y Y | A A | A A | A A | C C | C C | C C | C C | O O | O O | D D | D D | E E | E E | E E | E E | E E | I I | I I | I I | I I | G G | G G | I I | I I | H H | H H | J J | K K | L L | L L | L L | L L | L L | L L | N N | N N | N N | D D | O O | O O | O O | O O | O O | O O | R R | R R | I I | S S | S S | S S | S S | T T | T T | T T | T T | T T | U U | U U | U U | U U | U U | V V | V V | V V | V V | V V | W W | Y Y | Z Z | Z Z | Z Z | S S | S S | b b | C C | A A | F F | H H | K K | I I | N N | Q Q | O O | Q Q | P P | P P | T T | T T | U U | U U | Y Y | Z Z | E E | S S | P P | D D | D D | L L | L L | N N | N N | N N | A A | I I | O O | O O | O O | O O | O O | O O | I I | A A | A A | E E | I I | G G | O O | G G | K K | O O | O O | O O | O O

The left hand character should match the right hand one.
 y y | a a | A A | A A | C C | C C | C C | C C | O O | O O | D D | D D | E E | E E | E E | E E | E E | I I | I I | I I | I I | G G | G G | I I | I I | H H | H H | J J | K K | L L | L L | L L | L L | L L | L L | N N | N N | N N | N N | D D | O O | O O | O O | O O | O O | O O | R R | R R | I I | S S | S S | S S | S S | T T | T T | T T | T T | T T | U U | U U | U U | U U | U U | V V | V V | V V | V V | V V | W W | Y Y | Z Z | Z Z | Z Z | S S | S S | b b | C C | A A | F F | H H | K K | I I | N N | Q Q | O O | Q Q | P P | P P | T T | T T | U U | U U | Y Y | Z Z | E E | S S | P P | D D | D D | L L | L L | N N | N N | N N | A A | I I | O O | O O | O O | O O | O O | O O | I I | A A | A A | E E | I I | G G | O O | G G | K K | O O | O O | O O | O O

General case: Latin
 The css says: text-transform: uppercase.
 The right hand character should match the left hand one.
 Y Y | A A | A A | A A | C C | C C | C C | C C | O O | O O | D D | D D | E E | E E | E E | E E | E E | I I | I I | I I | I I | G G | G G | I I | I I | H H | H H | J J | K K | L L | L L | L L | L L | L L | L L | N N | N N | N N | N N | D D | O O | O O | O O | O O | O O | O O | R R | R R | I I | I I | I I | I I | U U | U U | V V | J J | K K | L L | L L | L L | L L | L L | L L | N N | N N | N N | D D | O O | O O | O O | O O | O O | O O | R R | R R | I I | I I | I I | I I | U U | U U | U U | U U | U U | U U | V V | W W | Y Y | Z Z | E E | S S | P P | D D | D D | L L | L L | N N | N N | N N | A A | I I | O O | O O | O O | O O | O O | O O | I I | A A | A A | E E | I I | G G | O O | G G | K K | O O | O O | O O | O O

Tests on this page:
 Latin1
 General case: Latin
 Special case: Finnish
 General case: Greek
 General case: Cyrillic
 General case: Armenian
 General case: Russian
 General case: Circled Latin
 General case: Fullwidth Latin
 General case: Deleted

Copyright © 2005 W3C (MIT, ERCIM, Keio)

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 & presentationW3C[®]

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 51

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

```
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 52

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


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](#)


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

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.

Separating content & presentation

Some lessons from this section



- Don't use presentation-related tags or embedded styling in HTML, use CSS for presentation.
- Keep HTML simple and use it to describe meaning (semantic markup).
- When creating new content with scripting, use class or id attributes to identify places where styling is to be applied, don't hardcode the styling in your script.

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

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



Composite messages

W3C[®]

Composite messages

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

The slide features a blue header bar with the text "Composite messages" on the left and the W3C logo on the right. The main body of the slide is white with the text "Composite messages" centered. At the bottom, there is a small copyright notice on the left and "slide 55" on the right.

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。
تكلّمونها من بيتي الجديد

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

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 57

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 58

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[®] 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/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 s</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 ">".
- Info** Line 14 column 93: start tag was here.

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

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[®]

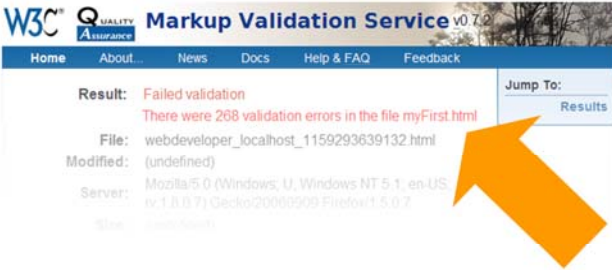
W3C[®] QUALITY ASSURANCE Markup Validation Service v0.7.0

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/20080509 Firefox/1.5.0.7
Size: 10000000

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 60

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

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: webdevelo...localhost_1159293639132.html
Modified: (undefined)
Server: Mozilla/5.0 (Windows NT 5.1; en-US; rv:1.8.8.7) Gecko/20070915 Firefox/1.5.0.7
Size: 12345678

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.";
```

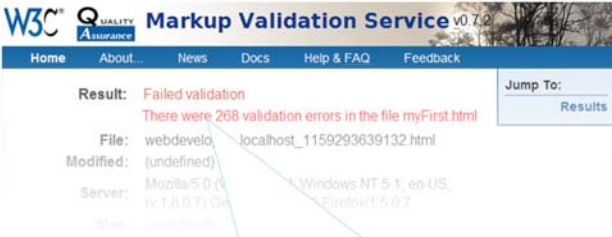
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 61

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[®]



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:1.8.8.7) Gecko/20070915 Firefox/1.5.0.7

Jump To: Results

Die Datei 268 enthält myFirst.html Gültigkeitsfehler. X

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

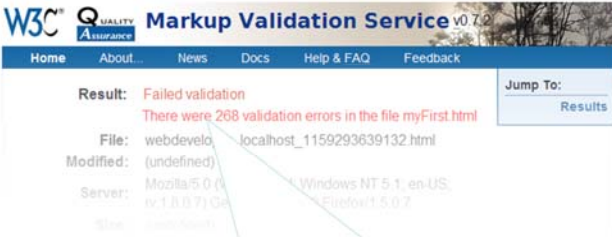
```
echo "Die Datei ".$errors." enthält ".$filename.  
" Gültigkeitsfehler.";
```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 62

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

W3C[®]
Composite messages

Handling variables



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 63

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: webdevelo...localhost_1159293639132.html
Modified: (undefined)
Server: Mozilla/5.0 (Windows NT 5.1; en-US; rv:1.8.8.7) Gecko/1.8.8.7 Firefox/1.5.0.7

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 64

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[®]

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

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 65

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[®]

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: webdev..._localhost_1159293639132.html
Modified: (undefined)
Server: Mozilla/5.0 (Windows NT 5.1; en-US; rv:1.8.0.13) Gecko/20080909 Firefox/1.5.0.7

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."
)
```

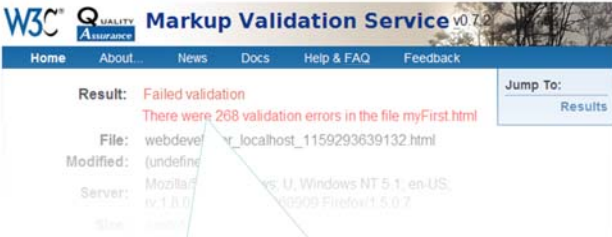
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 66

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.

W3C[®]
Composite messages

Handling variables



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

1 ошибка

2 ошибки

5 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 67

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 68

The diagram illustrates a sentence-like composite message. At the top, the text 'The < > is on.' is displayed. Below this, three words are arranged: 'printer' on the left, 'stapler options' in the center, and 'stacker' on the right. Three yellow arrows originate from these words and point upwards towards the angle brackets in the sentence above. The arrow from 'printer' points to the left angle bracket, the arrow from 'stapler options' points to the space between the angle brackets, and the arrow from 'stacker' points to the right angle bracket. This visualizes how the words are mapped to the placeholders in the sentence.

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

W3C[®]

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 69

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 70

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 71


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



- 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 72

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

What this means to me

Recommendations in conclusion

W3C[®]

- be aware that well localized deliverables and content begins with **design & development**, not the localization folks
- designers and developers need to find out about good internationalization practice
- localizability is often achieved by following standard good practices

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

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.

Getting more information
W3C can help


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

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 74

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

We could always do with help and support for this.



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

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