Objectives

You will be able to tell your friends and colleagues:

- Why localization is not just a question of grabbing a technical guy to translate stuff
- Why you need to think about localization earlier than people typically expect
- Insights into internationalization at the W3C
Overview

W3C's I18n Activity
L10n or i18n?
Content vs. presentation
I18n overview
  Characters
  Document formats
  Presentation matters
  Practical barriers
  Cultural differences
Summary
W3C Internationalization Activity
About the W3C

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

Keywords:
consensus and vendor neutrality.

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

W3C Internationalization Activity
About the W3C

MIT
ERCIM
Keio

3 Hosts
15 Offices
60 Staff (20 locations)
We are real (and very nice) people!
W3C Internationalization Activity
About the W3C

HP, Microsoft, Sun, WebMethods, Sony, Fujitsu, Software AG, IBM, Apple, Elisa, Nokia, Siemens, Vodafone, DoCoMo, T-Online, Academia Sinica, Fraunhofer, MIT, CSIRO, EUnet, ETRI, ERCIM, Boeing, ChevronTexaco, Agfa, DaimlerChrysler, Elsevier, and many more...

3 Hosts
14 Offices
60 Staff (20 locations)
400 Members
43 Liaisons

2005: New fee structure to assist developing countries!
Introduction to Internationalization

W3C Internationalization Activity

About the W3C

Architecture
- DOM
- XML
- Internationalization
- URI
- Web Services

Interaction
- Device Independence
- Graphics
- HTML
- Math
- Mobile Web
- Multimodal Interaction
- Rich Web
- Style
- Synchronized Multimedia
- Voice Browser
- XForms

Technology & Society
- Patent Policy
- Privacy
- Semantic Web

Web Accessibility
- International Program Office
- Technical Activity

Other
- QA
- Incubator
- Technical Architecture Group
- Advisory Board

4 Domains
23 Activities
39 Working Grp
19 Interest Grp
6 Coordination Grp
Internationalization Activity

Core Working Group
Reviews, advice, and internationalization specifications

ITS (Internationalization Tag Set) Working Group
Elements and attributes for schema developers

GEO (Guidelines, Education & Outreach) Working Group
Making internationalization aspects of W3C technology better understood and more widely and consistently used
Introduction to Internationalization

Overview

W3C's I18n Activity
L10n or i18n?
Content vs. presentation
I18n overview
   Characters
   Document formats
   Presentation matters
   Practical barriers
   Cultural differences
Summary
• "Visitors linger twice as long as they do at English-only URL’s.

• Business users are 3 times more likely to buy when addressed in their language.

• Customer service costs drop when instructions are displayed in the user's native language."

'Strategies for Global Sites'
Donald DePalma
Forrester Research Inc.
"One large IT company discovered that a significant percentage of inquiries were coming from South Korea - they created a Korean website and revenues rose by 8 percent."

'Global eCommerce'
Donald J. Plumley
Bowen Global Solutions
Localization

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

Internationalization

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

http://www.w3.org/International/questions/qa-i18n

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

W3C's I18n Activity
L10n or I18n?
Content vs. presentation
I18n overview
- Characters
- Document formats
- Presentation matters
- Practical barriers
- Cultural differences

Summary
Introduction to Internationalization

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

To the right is some CSS code that applies styling to the HTML.
The HTML is shown on the left. There is no presentational information in the HTML — which is as it should be.

To the right is some CSS code that applies styling to the HTML.
Each of these windows shows EXACTLY the same HTML file. The changes made to the CSS file produced three very different presentations of that basic content.

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

This can also be beneficial for localization, since typographic approaches, colors, etc, may need to be changed for different locales. Making such changes in the CSS is much easier than adapting the HTML.
Remember, also, that the Mobile Web is becoming increasingly important these days – and may be especially so in developing countries in the future. This means that content needs to be adapted to fit on handheld devices with smaller screens.

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

You should not make assumptions, when creating content, that you know what it will look like when finally displayed. These days, it may well be displayed in a number of different formats.
Here are some ways in which typographic differences may appear between language versions of the same content.
Separating content & presentation

International issues

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

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.
Making the World Wide Web *worldwide*.

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

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

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

<p>See the <span class="bold">System Administrator Guide</span> for an example of re-use.</p>

The same applies to document conventions such as representation of referenced resources. When using class annotations or microformats, don't describe the expected presentational rendering, describe the function of the text.
See the **System Administrator Guide** for an example of re-use.

<p>See the <span class="doctitle">System Administrator Guide</span> for an example of re-use.</p>
Overview

W3C's I18n Activity
L10n or i18n?
Content vs. presentation
I18n overview
Characters
Document formats
Presentation matters
Practical barriers
Cultural differences
Summary
English is just another language.
This kind of multilingual text on a single page was very rare only 10 years ago.
Early character sets based on 7-bit bytes, gave $2^7$ (ie. 128) possible characters.
Adding an 8th bit gave a total of 256 possible characters. Still this was not enough for all European needs.
The code page mechanism, where the meaning of the upper cells was changed according to context helped a little, but was very messy.

It still didn't come close, however, to addressing the needs of the Far East, where the character sets had to incorporate thousands of ideographic characters at a time.
Unicode solves this problem.

It is a single character set that covers all the commonly used scripts of the world in one place. This allows for simple display and storage of multilingual content, and for easy transitions between localized content.
An 'encoding' refers to the way that characters are mapped from the character set to bytes in the computer. Different encodings yield different byte sequences.

To emphasize that character sets and encodings are different things, note how Unicode has three possible encodings, even though the actual character set is just defined once. In order to correctly interpret byte sequences and convert them into the right characters, you need to know what encoding was used.
You must declare the encoding of your content somewhere, so that it can always be discovered by any application that wants to interpret the text.

There are a number of ways of doing this. For more information see http://www.w3.org/International/tutorials/tutorial-char-enc/.

Note that you must also save your data in the appropriate encoding – labelling alone is not sufficient (see http://www.w3.org/International/questions/qa-changing-encoding).
In an encoding such as UTF-8 characters can be encoded using a mixture of 1 to 4 bytes. This means that when manipulating, comparing, pointing into, wrapping, or styling data, etc., you need to know where the character boundaries are, and never separate the bytes that constitute a single character.
This sequence of slides shows how a cursor would have to jump through the bytes in memory as you press the right cursor key.
I18n Overview: Characters
Working with characters

אָאָא ああ

61 D7 90 E3 81 82 61 D7 90 E3 81 82
Introduction to Internationalization

I18n Overview: Characters
Working with characters

אףあああ

61 D7 90 E3 81 82 61 D7 90 E3 81 82

Copyright © 2005 W3C (MIT, ERCIM, Keio)
You need to ensure that the applications you are dealing with – especially any back-end scripting – can appropriately deal with text.

This slide shows a photo uploaded to Flickr with XMP meta data in UTF-8. The Flickr user interface, which supports UTF-8, has taken the title of the photo from the XMP data, but some backend process has mangled the encoding. You can guess at the meaning of this title, but text in, say, Chinese, would be completely unreadable.

Be careful that the functions you use in languages such as PHP and Python can handle multibyte characters correctly, and that encoding information is recognized and appropriately dealt with.
If you are running processes on text, you may also want to normalize the text beforehand to make it easier to collate character sequences in Unicode that are different but canonically equivalent.
There is a lot of demand for people to be able to use non-ASCII characters in Web addresses.
New standards have come out of the IETF recently that make this possible. The W3C personnel contributed to the development of these standards.

There are still some hurdles to overcome with regard to security and deployment, but it is possible to use these now. For more information see http://www.w3.org/International/articles/idn-and-iri/.
Introduction to Internationalization

Overview

W3C's I18n Activity
L10n or i18n?
Content vs. presentation
I18n overview
  Characters
  Document formats
  Presentation matters
  Practical barriers
  Cultural differences
Summary
Applications exist that can use natural language information about content to deliver to users the most relevant information or styling according to their language preferences. The more content is tagged and tagged correctly, the more useful and pervasive such applications will become.

There are a number of possible ways to declare language information in HTML, but the effectiveness and the rules that apply to each approach vary. For more information see http://www.w3.org/TR/i18n-html-tech-lang/.
In particular, it is important to recognize that there are two different types of language declaration. Different mechanisms (shown on the previous page) naturally fall into one or other of the different types.
In addition to language declarations, there are other types of markup that are needed to support non-Latin scripts. One important example is markup to support bidirectional text in languages based on Arabic or Hebrew scripts.

If you develop content for these languages, you must become familiar with their use (see for example http://www.w3.org/International/articles/inline-bidi-markup/). If you develop schemas, you should ensure that you provide such constructs for others to use.

The ITS (International Tag Set) Working Group at the W3C is currently specifying markup that can be used to support international use of documents, and also efficient localization of documents.
In addition to language declarations, there are other types of markup that are needed to support non-Latin scripts. One important example is markup to support bidirectional text in languages based on Arabic or Hebrew scripts.

If you develop content for these languages, you must become familiar with their use (see for example http://www.w3.org/International/articles/inline-bidi-markup/). If you develop schemas, you should ensure that you provide such constructs for others to use.

The ITS (International Tag Set) Working Group at the W3C is currently specifying markup that can be used to support international use of documents, and also efficient localization of documents.
In addition to language declarations, there are other types of markup that are needed to support non-Latin scripts. One important example is markup to support bidirectional text in languages based on Arabic or Hebrew scripts.

If you develop content for these languages, you must become familiar with their use (see for example http://www.w3.org/International/articles/inline-bidi-markup/). If you develop schemas, you should ensure that you provide such constructs for others to use.

The ITS (International Tag Set) Working Group at the W3C is currently specifying markup that can be used to support international use of documents, and also efficient localization of documents.
At the VISTA console, submit a job to print. (Refer to “Submitting a Job” in Chapter 5.)

At the operator control panel, make sure the printing system is in Make-Ready mode. The MAKE-READY/RUN indicator should not be lit.

Press the START button to sound the horn. The MAKE READY / RUN indicator flashes.

At the third beep, press the START button again. The START indicator remains lit and paper movement begins.

When the web reaches minimum print speed, the test pattern prints.

Press the MAKE-READY/RUN button to place the printing system in Run mode and start printing the live test pages. The MAKE-READY/ RUN indicator should be lit.

An example of markup that can help make translation more efficient is the provision of a flag to indicate whether or not text should be translated. This can be used by translation tools to screen text from translators or machine translation systems where necessary.

In this example of product documentation, 'START' and 'MAKE-READY/RUN' appear on a hard panel that will not be translated. The markup can be used to indicate that. In actuality, the ITS group will come up with a number of ways of implementing a translate flag. In some cases these may be used by content authors, in other cases they may be applied via rules. For more detail, follow the development of the working draft at http://www.w3.org/TR/its/.
Von der VISTA-Konsole aus einem Druckauftrag übermitteln. (Siehe hierzu “Auftrag übergeben” in Kapitel 5.)

Am Steuerpult prüfen, ob der Make-Ready-Modus aktiv ist. (Die Anzeige MAKE-READY/RUN darf nicht leuchten).

START drücken, so dass die Hupe ertönt und die Anzeige MAKE READY / RUN blinkt.

Beim dritten Ton erneut START drücken. Die Anzeige START leuchtet konstant und der Papiertransport läuft an.

Von der VISTA-Konsole aus einen Druckauftrag übermitteln. (Siehe hierzu “Auftrag übergeben” in Kapitel 5.)

Am Steuerpult prüfen, ob der Make-Ready-Modus aktiv ist. (Die Anzeige MAKE-READY/RUN darf nicht leuchten).

START drücken, so dass die Hupe ertönt und die Anzeige MAKE READY / RUN blinkt.

Beim dritten Ton erneut START drücken. Die Anzeige START leuchtet konstant und der Papiertransport läuft an.

Press the \(<span translate="no">START</span>\) button to sound the horn. The \(<span translate="no">MAKE-READY/ RUN</span>\) indicator flashes.
Volcanic eruptions have literally devastated large inhabited areas. During the 1914 eruption of Sakurajima in Kyushu, 687 houses in Kurokami were buried in hot ash. What remained of this shrine gate, previously five meters tall, was left as a reminder.

Can't mark up for language, bidirectional markup, abbreviation, styling, etc.

In some cases, an approach to schema design is important, rather than specific tags. For example, the Japanese text in an attribute value shown here cannot be marked up for language, directionality, abbreviation, styling, etc, since it is part of the attribute text.
Volcanic eruptions have literally devastated large inhabited areas. During the 1914 eruption of Sakurajima in Kyushu, 687 houses in Kurokami were buried in hot ash. What remained of this shrine gate, previously five meters tall, was left as a reminder. 

<image src="kk-torii.jpg" height="180" width="240">
  <caption>Kurokami maibutsu gate (腹五社神社黒神埋没鳥居), Sakurajima Island.</caption>
</image>

It would have made more sense to use an element for the caption.

The ITS Working Group will also provide advice of this kind to schema developers.

The I18n Core Working Group has also discussed concepts such as this with other W3C working groups. For example, XHTML 2 will hopefully address a number of situations in HTML where text cannot be marked up appropriately.
Overview

W3C's I18n Activity
L10n or i18n?
Content vs. presentation
I18n overview
  - Characters
  - Document formats
  - Presentation matters
  - Practical barriers
  - Cultural differences
Summary
The term 'glyph' refers to the visual representation of an underlying character.

There is only one character code for the Latin letter a, but the visual representation can have many different forms, depending on font, emphasis, etc.
Arabic letters join up with adjacent letters, so there can be up to four different forms for each letter when displayed. You use the same character code, however, for a given letter of the alphabet, and the different glyph shapes are chosen from the font according to the position of the letter at run time.
This slide shows how characters can be visually represented in apparently different orders and with different shapes in Hindi, also.

Unicode is making a similar distinction between content and presentation as we saw earlier. This enables easier text analysis and comparison, given that the underlying codes represent semantic rather than presentational information.
CSS3 holds the promise of a number of typographic approaches that are needed for non-Latin scripts, such as Chinese and Japanese.
A mechanism for support of vertical text is being proposed. Discussions are currently centered on mixing text of different directionalities on the same line.

This and the following slide indicate some of the changes that are applied to text as its orientation is changed. Note that this should just be presentational sugar. There should be no need to alter the content, just the styling.
When these new typographic features are available and supported in user agents, developers and content authors will need to familiarize themselves with the numerous properties that are available.

Before that, if you use a non-Latin script, you should check that your requirements have been taken into account. This slide shows a picture of vertical text on an Indian doorway that I came across recently. We will need to check that the vertical text properties in CSS take into account that the text proceeds downwards syllable by syllable, not letter by letter.
This and the following slide illustrate how different scripts exhibit different wrapping behavior at the end of a line.

It is important to ensure that user agents perform such wrapping correctly. It is also important to ensure that all the user parameters that are needed to control wrapping are available to the styling mechanism (e.g. CSS).
### I18n Overview: Presentation matters

**Script-specific typography**

Richard Ishida, Consultant, Design of International User Interfaces

<table>
<thead>
<tr>
<th>Language</th>
<th>Script-specific Typography</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>global</td>
</tr>
<tr>
<td>Arabic</td>
<td>伸びる （123-456）</td>
</tr>
<tr>
<td>Japanese</td>
<td>設計 （123-456）</td>
</tr>
</tbody>
</table>

[http://people.w3.org/rishida/scripts/samples/wrapping.html](http://people.w3.org/rishida/scripts/samples/wrapping.html)
As this and the next slide show, Arabic justification stretches words rather than spaces. Another example of script-differentiated behaviour.
I18n Overview: Presentation matters
Script-specific typography

والدهان والمخرج والتاجير
والضابط والحي والميت، ويومئ
اليناء فو از حديد عوسم ۹۰۹ اذ
يهيئ حسن فكرت عشر
مسرحيات، ثم يومئ اذ تراود
Because the predominant reading order for Arabic and Hebrew script-based text is right-to-left, the layout of pages also tends to be a mirror-image of a page in a European language.
Icons may also have directional bias. In some cases the icon can simply be flipped horizontally, in others a new icon will need to be created.

Tables, spreadsheets, charts, collated pictures, etc. also tend to flow from right to left where a right to left script is used.
This slide provides some examples of differences between English and Arabic approaches to mathematical presentation. The W3C has recently produced a note about this, with a view to enabling the various Arabic approaches in the future.

We are always looking out for other requirements, related to non-Latin typography. If you are aware of things that the Web should support, please let us know.

This section on presentation invites you to:
- find out and use features that are currently available
- design your applications in an extensible way, so that these features can be incorporated when needed for international content
- push for new features to be implemented by user agents – getting support in the W3C standards is not sufficient, the user agent developers must also be convinced that they should support them – this means both pushing for feature to be supported, and using them when they are made available.
A recent workshop in Beijing explored international requirements for markup to support speech synthesis. There are plans to organize another workshop soon.

Since there are no spaces between words in Chinese, the sentence above can be read in a number of different ways. Markup to show word boundaries when needed for disambiguation was one of the results of the Beijing workshop.
Overview

W3C's I18n Activity
L10n or i18n?
Content vs. presentation
I18n overview
Characters
Document formats
Presentation matters
Practical barriers
Cultural differences
Summary
Tokyo hotel:
- Cooles and heates. If you want the condition of cool in your room, please control yourself.

Paris hotel:
- Guests are expected to complain at the office between the hours of 9am and 10am each day.

Tokyo hotel:
- You are invited to take advantage of the chambermaid.

Vienna hotel:
- In case of fire, do your utmost to alarm the porter.

This slide helps make the point that translation is not simply word substitution, but that there can be complex and subtle factors to take into account to achieve a good rendering of the meaning.
They are speaking to her from my new house.

Están hablándole desde mi casa nueva.

私の新しい家から彼女と話しています。

تكمَّموها من بيتي الجديد

This slide shows the same idea expressed in multiple languages. Within each translation of the sentence, the number of words is different, and the order of those words changes.
I18n Overview: Practical barriers
Text fragmentation & re-use

There were %d spelling mistakes in file: %s.
Datei %s enthält %d Rechtschreibfehler.

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

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

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

This is an example of syntax differences affecting development techniques.

The order of variables needs to be different between English and German versions. Unless you are using slightly more advance techniques in PHP, you will prevent this possibility and seriously affect translatability.
In this example, the developer has tried to save memory by re-using part of a common sentence. Unfortunately, because of the effects of rules about agreement between gender and number in many languages, this becomes an untranslatable phrase. The developer needs to be aware of the likely impact on translatability of such things.
English and Chinese text usually expand when translated. You should consider the potential impact of this on page design, and either allow text to flow into larger areas, or leave expansion space.

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

W3C's I18n Activity
L10n or i18n?
Content vs. presentation

I18n overview
Characters
Document formats
Presentation matters
Practical barriers
Cultural differences

Summary
Be careful about assuming what others' name and address formats will be. Also think about how you will store the names and addresses in the database. For example, do you really need to split out street number? How will you generate a Russian or Japanese address that goes from general to specific from top to bottom?
Symbolism can differ from place to place. For example the check mark means *incorrect* in some places around the world.

Ensure that you do not give the wrong message through your use of colors, symbolism, examples, etc.
Symbolism can differ from place to place. For example the check mark means *incorrect* in some places around the world.
Graphics may need to be changed if they don't reflect the local culture of certain places.
Body language and gestures are particularly dangerous. Each of these symbols can give offense in one part of the world or another.
When dealing with graphics, consider how to deal with text. Ideally the text will be overlaid on a graphic, rather than embedded in it. If the text is within the graphic, try to ensure that you develop it in layers, with text on a separate layer, so that when it comes to translation the text can be easily removed and replaced over complicated backgrounds.
Be wary of humor. It doesn’t travel well.

Also, try to ensure that examples used in text are understandable by the audience of the translated version. Use examples from a variety of cultures where possible.
Color also has different connotations in different parts of the world.
For example, a black wedding kimono is not as strange in Japan as it may seem to a European.
Then you need to be aware that people in different parts of the world may do things in different ways. For example, the radar chart was such a common way of representing comparative data in Japan that, when Lotus 1-2-3 was launched in that area they had to reengineer it to add that.
"... one Latin American teacher recently complained to me that the US-manufactured and well-translated educational software currently being used in his country's primary schools presupposed 'solitary problem solvers', whereas his culture stressed collective problem-solving."


Considerations of this kind require you to make big decisions at the very start of the development phase about how to proceed. Otherwise you could waste a lot of time and energy producing something that doesn't meet your customer's needs.
This and the following slides show how Yahoo adapts its categorizations to reflect the preoccupations of various different countries.
I18n Overview: Cultural differences

Different approaches

<table>
<thead>
<tr>
<th>Actualités et médias</th>
<th>Sports et loisirs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journaux, Télévision, Média</td>
<td>Foot, Tourisme, Auto, Moto, Voile</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Commerce et économie</th>
<th>Art et culture</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2B, Shopping, Emploi, Immobilier</td>
<td>Littérature, Cinéma, Musique, BD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Informatique et Internet</th>
<th>Divertissement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet, Logiciels, Fonds d'écran</td>
<td>Tests/Quiz, Loterie, Humour, Sorties</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Santé</th>
<th>Classement géographique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diététique, Médecine, Thérapie</td>
<td>Pays, Europe, France, Paris</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enseignement et formation</th>
<th>Références et annuaires</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primaire, Secondaire, Supérieur</td>
<td>Dictionnaires, Annuaires, Cartes/Atlas</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Institutions et politique</th>
<th>Société</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ministres, Droit, Politique</td>
<td>Enfants, Gastronomie, Rencontres</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sciences et technologies</th>
<th>Sciences humaines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astronomie, Physique</td>
<td>Archéologie, Histoire, Psychologie</td>
</tr>
</tbody>
</table>
I18n Overview: Cultural differences
Different approaches

Yahoo!カテゴリー

エンターテインメント
映画,音楽,芸術,コミック,占い...

趣味とスポーツ
アウトドア,ゲーム,車,スポーツ...

芸術と文学
学問,建築,美術館,歴史,文学...

生活と文化
テレビ,環境,グルメ,警察...

情報
コンピュータとインターネット
ネットワーク,ソフトウェア,WWW...

政治
大学,専門学校,小中高,選挙...

健康と科学
病院,医療,ダイエット...

地域情報
日本の地方,世界的な観光地...

各種資料と情報源
図書館,辞書,時報,鳥類等 XVI

 diferente

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

W3C's I18n Activity
L10n or I18n?
Content vs. presentation
I18n overview
Characters
Document formats
Presentation matters
Practical barriers
Cultural differences

Summary
Summary
The value of internationalization

Localization can help bridge the Digital Divide

Localization can be a lot more than just translation

Treating localization as an afterthought only makes things worse

- Higher cost during localization due to rework and support activities
- Longer localization schedules, and slower time to market/release
- Reduced quality user perceptions, despite the extra effort
Internationalization means:

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

- ensure that the needs of your scripts and languages are brought to the attention of the W3C and other standards and tool developers
- push for adoption of international technology by user agents and content creation/management tools
- keep presentation and content separate
- always use standards and validate your code
- follow best practices for internationalization, accessibility, the mobile web, etc.
All the GEO materials are available from the Internationalization home page.

http://www.w3.org/International/
Introduction to Internationalization

<table>
<thead>
<tr>
<th>Summary</th>
<th>GEO resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td></td>
</tr>
<tr>
<td>- International &amp; multilingual Web sites</td>
<td></td>
</tr>
<tr>
<td>- Monolingual vs. multilingual Web sites</td>
<td></td>
</tr>
<tr>
<td>- Serving XHTML 1.0</td>
<td></td>
</tr>
<tr>
<td>Characters &amp; encoding</td>
<td></td>
</tr>
<tr>
<td>- Character encodings</td>
<td></td>
</tr>
<tr>
<td>- Document character set</td>
<td></td>
</tr>
<tr>
<td>- Who uses Unicode?</td>
<td></td>
</tr>
<tr>
<td>- Upgrading to Unicode</td>
<td></td>
</tr>
<tr>
<td>- An Introduction to Multilingual Web Addresses</td>
<td></td>
</tr>
<tr>
<td>- Language tags in HTML and XML</td>
<td></td>
</tr>
<tr>
<td>- The HTTP Charset parameter</td>
<td></td>
</tr>
<tr>
<td>- Checking HTTP headers</td>
<td></td>
</tr>
<tr>
<td>- Checking the character encoding using the validator</td>
<td></td>
</tr>
<tr>
<td>- CSS character encoding declarations</td>
<td></td>
</tr>
<tr>
<td>- Multilingual form encoding</td>
<td></td>
</tr>
<tr>
<td>- HTML, XHTML, XML and control codes</td>
<td></td>
</tr>
<tr>
<td>- Unexpected characters or blank lines</td>
<td></td>
</tr>
<tr>
<td>- Setting encoding in web authoring applications</td>
<td></td>
</tr>
<tr>
<td>- Setting <code>charset</code> information in .htaccess</td>
<td></td>
</tr>
<tr>
<td>- Missing characters and glyphs</td>
<td></td>
</tr>
<tr>
<td>- Non-English tags</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
</tr>
<tr>
<td>- Why use the language attribute?</td>
<td></td>
</tr>
<tr>
<td>- Using HTTP and meta for language information</td>
<td></td>
</tr>
<tr>
<td>- Language tags in HTML and XML</td>
<td></td>
</tr>
<tr>
<td>- 2-letter or 3-letter language codes</td>
<td></td>
</tr>
<tr>
<td>- Styling using the lang attribute</td>
<td></td>
</tr>
<tr>
<td>- Setting language preferences in a browser</td>
<td></td>
</tr>
<tr>
<td>- When to use language negotiation</td>
<td></td>
</tr>
<tr>
<td>- Apache MultiViews language negotiation set up</td>
<td></td>
</tr>
<tr>
<td>- Accept-Language used for locale setting</td>
<td></td>
</tr>
</tbody>
</table>

Bidirectional text
- What you need to know about the bidi algorithm and inline markup
- Bidi space loss
- XHTML/XML & bidi formatting codes vs. markup
- CSS vs. markup for bidi support

Resource identifiers
- An Introduction to Multilingual Web Addresses

Local formats
- Displaying formatted dates
- Select box sorting
- Dates and time

Miscellaneous
- Display capabilities
- Script direction & languages
- Introduction to Indic Scripts
- Ruby

Navigation
- Global gateway pull-downs

Other
- XSLT for multilingual output
- Using the link element

There are many articles available on the W3C Internationalization subsite (http://www.w3.org/International/). This shows a selection.
In addition, the GEO Working Group is developing techniques documents, to provide advice on accomplishing particular tasks.
We are trying to provide multiple points of entry and levels of detail to the techniques information.
As designers and developers:

- be aware that well localized deliverables and content begins with you, not the localization folks
- develop an awareness of good internationalization practice
- put in place a mechanism for obtaining user feedback on an international level
- don’t be locked into a short-term view with development costs – realize the value of enabling expansion beyond the local market/users
- always use standards (especially Unicode)
Get involved:

- visit the I18n Activity Home Page
- join a W3C Internationalization Working Group, or the Interest Group (www-international@w3.org)
- offer to help with reviews, or provide local knowledge for other Working Groups
- provide translations of W3C specifications or articles
- take advantage of the i18n-readiness of W3C technology
There are a number of well-accepted goals in development these days, expressed by words ending in -bility - interoperability, scalability, portability, accessibility. In my mind, in today's ever increasingly global community, another key -bility is localizability.