11.7 Internationalization

Dr. Richard Isida, W3C
### Characters, the foundation

<table>
<thead>
<tr>
<th>Standard</th>
<th>Country</th>
<th>No. of Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>JIS X 0208:1997</td>
<td>Japan</td>
<td>6,879</td>
</tr>
<tr>
<td>JIS X 0212:1990</td>
<td>Japan</td>
<td>6,067</td>
</tr>
<tr>
<td>KS X 1001:1992</td>
<td>Korea</td>
<td>8,224</td>
</tr>
<tr>
<td>GB 2312-80</td>
<td>China</td>
<td>7,445</td>
</tr>
<tr>
<td>Big Five</td>
<td>Taiwan</td>
<td>13,494</td>
</tr>
<tr>
<td>CNS 11643-1992</td>
<td>Taiwan</td>
<td>48,711</td>
</tr>
</tbody>
</table>

---

**European alphabetic scripts**
- Latin
- Greek
- Cyrillic
- Armenian
- Georgian
- Hunic
- Ughnam
- Modifier letters
- Combining characters

**East Asian scripts**
- Han
- Hiragana
- Katakana
- Hangul
- Bopomofo
- Yi

**Middle East scripts**
- Hebrew
- Arabic
- Syriac
- Thaana

**South & South East Asian scripts**
- Devanagari
- Bengali
- Gurmukhi
- Gujarati
- Panjabi
- Oriya
- Tamil
- Telugu
- Kannada
- Malayalam
- Sinhala
- Thai
- Lao
- Tibetan
- Myanmar
- Khmer

**Symbols**
- Currency symbols
- Letter-like symbols
- Mathematical operators
- Numeric forms
- Technical symbols
- Geometrical symbols
- Miscellaneous symbols & dingbats
- Enclosed & square brackets
- Braille

**Additional scripts**
- Ethiopic
- Cherokee
- Canadian Aboriginal
- Syllabics
- Mongolian

**Etc...**
Characters, the foundation

<table>
<thead>
<tr>
<th>Character</th>
<th>Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>41</td>
</tr>
<tr>
<td>á</td>
<td>C3 A1</td>
</tr>
<tr>
<td>a</td>
<td>E3 81 82</td>
</tr>
<tr>
<td>अ</td>
<td>F0 A3 8E B4</td>
</tr>
</tbody>
</table>

61 D7 90 E3 81 82 61 D7 90 E3 81 82
Characters, the foundation

日本語の文字列

Copyright © 2003 W3C (MIT, ERCIM, Keio).
Slide 11

Characters, the foundation

Copyright © 2003 W3C (MIT, ERCIM, Keio).
Slide 12
Internationalized Resource Identifiers

http://raksmorgas.josefsson.org/målfranzen.html

http://räksmörgås.josefsson.org/målfranzén.html

Easier to create
- ... memorize
- ... transcribe
- ... interpret
- ... guess / find things
- ... relate to (branding)

domain name path

http://raksmorgas-Swalo.josefsson.org/m%C3%A5l/franz%C3%A9n.html

- Phishing (www.paypal.com)
Internationalized document formats

The title says "W3C" in Hebrew.

Using the bidi algorithm only
The title says "W3C" in Hebrew.

Markup (characters ordered as in memory!)
"<span dir="rtl">W3C</span>"
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.

Kurokami mibutsu gate (倉吉神社黒木堂) became a reminder of the eruption. Sakurajima Island.
Presentation Matters

国際サッカー連盟（FIFA）本部での理事
会を前に記念撮影におき
まる理事ら。前列右
から2人目がヨハンソ
ン（Johansson）
欧州サッカー連盟会長。

Copyright © 2005 W3C (W3C, World Wide Web Consortium)
Presentation Matters

http://people.w3.org/ishida/scripts/samples/wrapping.html

Copyright © 2005 W3C (MIT, ERCIM, Keio)
Presentation Matters

والدهان والمخرج والتـاجـر
والضاـبـط والـحـي والمـيت، ويومي
الـبـنـافـوـز حـديد عـوسـم 49 ًذ
بهـيـبـي حـسن فـكرت عــشـر
مسروـحـاـت، ثم يومي ًذ تراوـد

Supporting cultural differences

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

Kenneth Kachotz
Language International, May 1995
Enabling localization

Tokyo host:
- Coffee and beers: If you want the condition of cool in your room, please control yourself.

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

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

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

Enabling localization

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。

تكلموها من بيتي الجديد
Enabling localization

There were \(<X>\) spelling mistakes in file: \(<Y>\)

Datei \(<Y>\) enthält \(<X>\) Rechtschreibfehler.

The \(< >\) has been disabled.

printer  stacker

stapler options
Current Areas of Activity

How do I ...

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

Authoring Techniques for XHTML & HTML
Internationalization: Characters and Encodings 1.0

Abstract

This document provides techniques for authoring content in XML, XHTML, and HTML. These techniques are intended for use in authoring applications that use the character sets and encodings specified in this document. The techniques described in this document are based on the principles of ISO/IEC 10646[29] and the Unicode Standard[28].

Character Sets, Character Encodings, and Entities

Choosing a Page Encoding

Languages

The character sets, character encodings, and entities defined in this document are intended for use in authoring applications that use the character sets and encodings specified in this document. The techniques described in this document are based on the principles of ISO/IEC 10646[29] and the Unicode Standard[28].

Specifying a Page Encoding

1. Choose a page encoding. 2. Prepare your documentation.

1. Choose a page encoding. 2. Prepare your documentation.

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

http://www.w3.org/International/
Getting Involved

- ensure that the needs of your scripts and languages are brought to the attention of the W3C developers
- always use standards and validate your code
- push for adoption of international technology by user agents
- editing environments are also key to success
- importance of the mobile web

Getting Involved

- visit the I18n Activity Home Page
- join a Task Force, the Interest Group, or the public mailing list
- offer to help with reviews, or provide local knowledge for other WGs
- provide translations of W3C specifications or articles
- take advantage of the i18n-readiness of W3C technology
Current Areas of Activity

Internationalization Activity

Core Working Group

ITS (Internationalization Tag Set) Working Group

GEO (Guidelines, Education & Outreach) Working Group

Getting Involved

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

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