



## Internationalizing Web Content

Richard Ishida  
W3C Internationalization Lead

## Objectives W3C

- Explore the dimensions of internationalization (i18n)
- Tease apart some basic contexts where internationalization is necessary
- Show examples of how the W3C is making local access to the Web easier/possible
- Show how internationalization is a prerequisite for good local content

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

## Overview W3C

- L10n or i18n?
- Getting the character basics right
- Extending technology to support local needs
- Removing barriers to international use
- Assessing cultural influences
- Improving the process
- Summary

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

## L10n or i18n? W3C

### 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/qe-i18n>

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

## Overview W3C

- L10n or i18n?
- Getting the character basics right
- Extending technology to support local needs
- Removing barriers to international use
- Assessing cultural influences
- Improving the process

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

## Getting the character basics right W3C

جعل شبكة الويب العالمية عالمية حقاً !

締造真正全球通行的万维网  
締造真正全球通行的萬維網

የአለም አቀፍን ድር ስልጠናት ለአለም አቀፍ ግድግዳ!

Κάνοντας τον Παγκόσμιο Ιστό πραγματικά Παγκόσμιο

יצר מהרשת רשת כלל עולמית באמת

वर्ल्ड वाईड वेब को सचमुच विश्वव्यापी बना रहे हैं !

Сделай Всемирную Паутину действительно всемирной!

Making the World Wide Web world wide!

ワールド・ワイド・ウェブを世界中に広げましょう

Hogy a Világháló valóban az egész világgá lehessen!

वर्ल्ड वाईड वेबलाई यथार्थमे विश्वव्यापी बनाउने !

"Дүниежүзілік торды" нағыз дүниежүзілік етеміз!

전 세계의 월드 와이드 웹으로 만들기!

हर सब वाईड वेब ठूँ वाक्यी हिमल-हिमली बनलए !

Сделаем "Всемирную паутину" действительно всемирной!

ကမ္ဘာတိုက် World Wide Web ပြုစုရမည့်အားဖြင့်

U ita uri Webu Nyangaredzi ya Dzhangho i vhe nyangaredzi ngangohol!

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

Getting the character basics right W3C

- European alphabetic scripts
  - Latin
  - Greek
  - Cyrillic
  - Armenian
  - Georgian
  - Runic
  - Ogham
  - Modifier letters
  - Combining characters
- East Asian scripts
  - Han
  - Hiragana
  - Katakana
  - Hangul
  - Bopomofo
  - Yi
- Middle East scripts
  - Hebrew
  - Arabic
  - Syriac
  - Thaana
- Symbols
  - Currency symbols
  - Letter like symbols
  - Mathematic operators
  - Numeric forms
  - Technical symbols
  - Geometrical symbols
  - Miscellaneous symbols & dingbats
  - Enclosed & square
  - Braille
- South & South East Asian scripts
  - Devanagari
  - Bengali
  - Gurmukhi
  - Gujurati
  - Punjabi
  - Oriya
  - Tamil
  - Telugu
  - Kannada
  - Malayalam
  - Sinhala
  - Thai
  - Lao
  - Tibetan
  - Myanmar
  - Khmer
- Additional scripts
  - Ethiopic
  - Cherokee
  - Canadian Aboriginal
  - Syllabics
  - Mongolian
- Etc....

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

Getting the character basics right W3C

|            |    |     |      |       |
|------------|----|-----|------|-------|
|            | A  | κ   | 好    | 丕     |
| Code point | 41 | 5D0 | 597D | 233B4 |

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

Getting the character basics right W3C

|            |        |             |             |                         |
|------------|--------|-------------|-------------|-------------------------|
|            | A      | κ           | 好           | 丕                       |
| Code point | 41     | 5D0         | 597D        | 233B4                   |
| Encodings  | UTF-8  | 41          | D7 90       | E5 A5 BD F0 A3 8E B4    |
|            | UTF-16 | 00 41       | 05 D0       | 59 7D D8 4C DF B4       |
|            | UTF-32 | 00 00 00 41 | 00 00 05 D0 | 00 00 59 7D 00 02 33 B4 |

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

Getting the character basics right W3C

Character *a* *a*  
 vs.  
 Glyph 雪 雪

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

Getting the character basics right W3C

وسيجمع ع

عندما يرد العالم أن يتكلم، فهو يتحدث لغة وينيكود. تسجل الآن حضور المؤتمر الدولي العاشر لونيكوند (Unicode Conference)، الذي سيعقد في 10-12 آذار 2007 بمدينة مالمو، ألمانيا. وسيجمع المؤتمر بين خبراء من كافة قطاعات الصناعة على الشبكة العالمية للتزيت ونيكوند، حيث ستمتد على الصعيدين الدولي والمحلي على حد سواء مناقشة سبل استخدام يونيكود في النظم القائمة وفيما يخص التطبيقات الحاسوبية، الخطوط، تصميم النصوص والخوسية بتعدد اللغات.

متعددة على

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

Getting the character basics right W3C

ह + ि + न + ् + द + ी

हिन्दी

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 12

Participating stakeholders:

- W3C adopted and monitors the use of Unicode as the document character set for its specifications, and liaises with the Unicode Consortium
- Unicode Consortium defines Unicode and expected character-level behaviors
- platform developers need to provide for Unicode support, such as rendering algorithms
- application developers should ensure that expected Unicode behaviors are implemented
- content developers and managers should use Unicode whenever possible

| Character | Bytes       |
|-----------|-------------|
| A         | 41          |
| á         | C3 A1       |
| あ         | E3 81 82    |
| 𐄌         | F0 A3 8E B4 |

a|x あ|x あ

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



a|x| あ|x| あ

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



a|x| あ|x| あ

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



NFC Ízelítőül  
 NFD Ízeliótoóüü

Ha a világ beszélni akarna, Unicode-*ul* szólna meg. Regisztráljon már most a Tizedik Nemzetközi Unicode Konferenciára, melyet 1997. március 10-12-én rendeznek Mainz-ban, Németországban. Ezen a konferencián az iparág több neves szakértője is részt vesz. *Ízelítőül* a témakból: a világháló és a Unicode nemzetköziesítése és lokalizálása, a Unicode alkalmazása működő rendszerekben és alkalmazásokban, szövegelrendezésnél, és többnyelvű számítógépeken.

## Getting the character basics right



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

## Getting the character basics right



```
<meta http-equiv="Content-type" content="text/html;charset=UTF-8" />
```

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

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

|                   | HTTP | <?xml .. | <meta .. |
|-------------------|------|----------|----------|
| HTML              | (✓)  | x        | ✓        |
| XHTML (text/html) | (✓)  | (✓)      | ✓        |
| XHTML (XML)       | (✓)  | ✓        | x        |

<http://www.w3.org/International/tutorials/tutorial-char-enc/>

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

## Overview



- L10n or i18n?
- Getting the character basics right
- Extending technology to support local needs
- Removing barriers to international use
- Assessing cultural influences
- Improving the process

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

## Extending technology to support local needs



Characters as ordered in memory:  
The title says "<span>פועילות הבינאום, W3C</span>" in Hebrew.

The title says "פועילות הבינאום, W3C" in Hebrew.

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

## Extending technology to support local needs



Characters as ordered in memory:  
The title says "<span>פועילות הבינאום, W3C</span>" in Hebrew.

The title says "פועילות הבינאום, W3C" in Hebrew.

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

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

## Extending technology to support local needs



Characters as ordered in memory:  
The title says "<span dir="rtl">פועילות הבינאום, W3C</span>" in Hebrew.

The title says "פועילות הבינאום, W3C" in Hebrew.

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

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

かみしばい  
これは紙芝居です。

```
<ruby>
  <rb>紙芝居</rb>
  <rt>かみしばい</rt>
</ruby>
```

<p>これは<ruby><rb>紙芝居</rb><rt>かみしばい</rt></ruby>です。</p>

這一晚會如常舉行

|              |                                           |
|--------------|-------------------------------------------|
| 這一 晚會 如常 舉行  | This banquet is held as usual.            |
| 這一 晚會 如 常 舉行 | If this banquet is held frequently.       |
| 這一晚 會 如常 舉行  | (An event) will be held tonight as usual. |

$$f(x) = \begin{cases} \sum_{i=1}^x x^i & \text{if } x < 0 \\ \int_1^x x^i dx & \text{if } x \in S \\ \tan \pi & \text{otherwise (with } \pi \approx 3.141) \end{cases}$$

ب إذا كان س .>  
ب إذا كان س >= م  
غير ذلك (مع π ≈ 3.141)

|                  |                |                    |
|------------------|----------------|--------------------|
| punctuation trim | 经验分<br>(万维)    | 经验分<br>(万维)        |
| auto-space       | 第10回のUnicode会議 | 第 10 回の Unicode 会議 |
| emphasis         | これは日本語の文章です。   | これは日本語の文章です。       |

当世界需要沟通时，请用 Unicode。将于3月10日-12日在德国 Mainz 市举行的第十届统一码国际研讨会在开始注册。本次会议将汇集各方面的专家。涉及的领域包括：国际互联网和统一码，国际化和本地化，统一码在操作系统和应用软件中的实现，字型，文本格式以及多文种计算等。

כאשר העולם רוצה לדבר, הוא מדבר ב־Unicode. הירשמו כעת לכנס Unicode הבינלאומי העשירי, שייערך בין התאריכים 10-12 במרץ, ב־מְיִינֶץ שבגרמניה. בכנס ישתתפו מומחים מכל ענפי התעשייה בנושא האינטרנט העולמי וה־Unicode, בהתאמה לשוק הבינלאומי והמקומי, ביישום Unicode במערכות הפעלה וביישומים, בגופנים, בפריסת טקסט ובמחשוב רב־לשוני.

Extending technology to support local needs



Richard Ishida Consultant, Design of International User Interfaces  
 ريتشارد ايشيدا - مستشار، تصميم عالمي لواجهات (global design) (456-123) التصميم للنص  
 ריצ'רד אישידה - יועץ, תכנון מיושם למשתמש 123-456 בין-לאומי  
 理查德·伊希达、国际化软件界面设计顾问、国际化软件界面、设计顾问「国际化软件界面设计顾问」  
 石田リチャード ユーザーインターフェース「国際化対応設計」コンサルタント。  
 리차드 이시다 국제어 인터페이스 기술컨설턴트  
 Ричард Ишида Консультант по интернационализации интерфейса пользователя  
 रिचर्ड ईशिदा परामर्शदाता, डिजाइन और इंटरनेशनल यूजर इंटरफेस

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

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 31

Extending technology to support local needs



Richard Ishida Consultant, Design of International User Interfaces  
 ريتشارد ايشيدا - مستشار، تصميم عالمي لواجهات (global design) (456-123) التصميم للنص  
 ריצ'רד אישידה - יועץ, תכנון מיושם למשתמש 123-456 בין-לאומי  
 理查德·伊希达、国际化软件界面设计顾问、国际化软件界面、设计顾问「国际化软件界面设计顾问」  
 石田リチャード ユーザーインターフェース「国際化対応設計」コンサルタント。  
 리차드 이시다 국제어 인터페이스 기술컨설턴트  
 Ричард Ишида Консультант по интернационализации интерфейса пользователя  
 रिचर्ड ईशिदा परामर्शदाता, डिजाइन और इंटरनेशनल यूजर इंटरफेस

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

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 32

Extending technology to support local needs



والدهان والمخرج والتاجر  
 والضابط والحي والميت، ويومئ  
 الينا فو از حديد عوسم ٤٩ اذ  
 يهبي حسن فكرت عشر  
 مسرحيات، ثم يومئ اذ تراود

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 33

Extending technology to support local needs



والدهان والمخرج والتاجر  
 والضابط والحي والميت، ويومئ  
 الينا فو از حديد عوسم ٤٩ اذ  
 يهبي حسن فكرت عشر  
 مسرحيات، ثم يومئ اذ تراود

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 34

Overview



- L10n or i18n?
- Getting the character basics right
- Extending technology to support local needs
- Removing barriers to international use
- Assessing cultural influences
- Improving the process

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 35

Removing barriers to international use



You 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。  
 تكلمونها من بيتي الجديد

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 36





Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 43



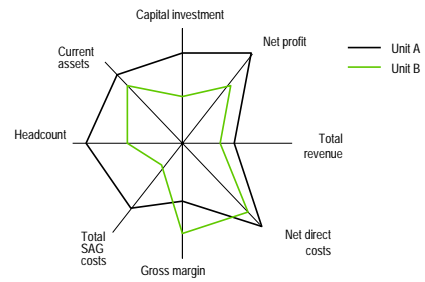
Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 44



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 45



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 46

"... 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 Kerislon,  
Language International, May 1996

Copyright © 2005 W3C (MIT, ERCIM, Keio)

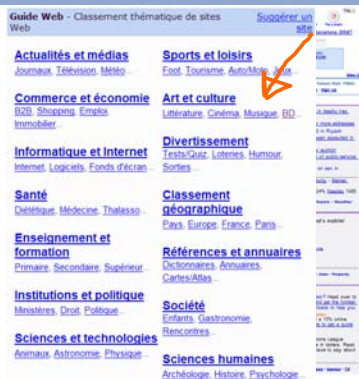
slide 47



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 48

## Assessing cultural influences



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 49

## Assessing cultural influences



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 50

## Overview

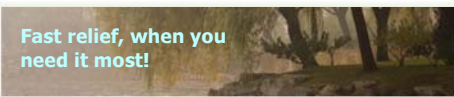


L10n or i18n?  
Getting the character basics right  
Extending technology to support local needs  
Removing barriers to international use  
Assessing cultural influences  
Improving the process

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 51

## Improving the process



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 52

## Improving the process

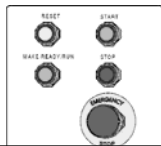


At the VISTA console, submit a job to print. (Refer to "Submitting a Job" in Chapter 5.)

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 flashes and the process



```
<para>
Press the
<span translate="no">START</span>
button to sound the horn. The
<span translate="no">MAKE-READY/ RUN</span>
indicator flashes.
</para>
```

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 55

## Improving the process

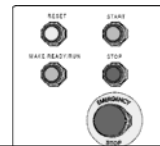


Von der VISTA-Konsole aus einem Druckauftrag übermitteln. (Siehe hierzu "Auftrag übergeben" in Kapitel 5.)

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 und der Prozess



```
<para>
Press the
<span translate="no">START</span>
button to sound the horn. The
<span translate="no">MAKE-READY/ RUN</span>
indicator flashes.
</para>
```

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 56

Content (XHTML)

```

<!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"
lang="zh-Hans">国际化活动万维网联盟</div>
<div class="description">
<p>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>
<p>The Activity comprises three Working Groups: Core, GEO
(Guidelines, Education & Outreach), and ITS (Internationalization
Tag Set). There is also an Internationalization Interest Group.</p>
</div>
</body>
</html>
    
```



Content (XHTML)

```

<!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"
lang="zh-Hans">国际化活动万维网联盟</div>
<div class="description">
<p>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>
<p>The Activity comprises three Working Groups: Core, GEO
(Guidelines, Education & Outreach), and ITS (Internationalization
Tag Set). There is also an Internationalization Interest Group.</p>
</div>
</body>
</html>
    
```

Presentation (CSS)

```

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;
}
    
```



- problems of resolution to support bold and italics in small CJK characters on-screen
- different ways of emphasizing text in Japanese (wakiten & amikake)

これは日本語です。  
これは日本語です。

- 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*.

<p>Making the World Wide Web ~~is~~ ~~worldwide~~ ~~is~~</p>

Making the World Wide Web *worldwide*.

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



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

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

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>



doctitle  
chaptertitle  
inputsequence  
etc.

L10n or i18n?  
Getting the character basics right  
Extending technology to support local needs  
Removing barriers to international use  
Assessing cultural influences  
Improving the process  
Summary

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)

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.

- i18n is more than just translation – it relates to enabling technologies to support local needs
- different groups of people work on different aspects of internationalization
- W3C drives fundamental Web-related technologies (but also works with other standards organizations)
- application and platform developers have to be encouraged to implement standards but also to enable other aspects of internationalization support
- content developers should use standards and observe best practices

- ◆ 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/>