

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.



Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

RESUMEN:

Introducción al variado y complejo tema de la internacionalización web, con ejemplos de problemas y posibles soluciones, e intentando demostrar que para acercarnos a otra cultura desde un sitio web no es suficiente con traducirlo.

Introducción

Leemos en los diarios títulos como los siguientes:

- "Rusia declara defender los derechos lingüísticos de los ucranianos rusohablantes".
- "El guarani pasa a ser uno de los idiomas oficiales del Mercosur".
- "El Parlasur aprobó el guarani como lengua de trabajo".
- "Estados Unidos: La policía no evita un asesinato múltiple por no traducir denuncia hecha en español"... y la lista de titulares sigue.

Pese a que otras noticias sigan ocupando más espacio en los noticieros, la diversidad lingüística y cultural surge cada vez con más importancia en la agenda diaria.

Nos guste o no, vivimos inmersos en algo llamado globalización o mundialización, intentando aprovechar sus beneficios y también sufriendo sus consecuencias. El mundo ha estado cambiando de bipolar o unipolar a multipolar, y esto nos está afectando cada vez más, en mayor o menor medida. Pero esto no pasa solamente en el ambiente de grandes negocios o política internacionales, sino que también un amplio abanico de personas (usuarios de redes sociales, compradores comunes, microemprendimientos y un largo etcétera) es incluido en esta marea incontenible.

Y esta marea la que nos hace cada vez más conscientes de la unidad pero, al mismo tiempo, de la diversidad de la sociedad mundial. Hoy podremos estar chateando con un amigo de China, mañana estar proponiéndole un negocio a una persona en Polonia y pasado mañana estamos intercambiando fotografías con alguien en una red social. Y conviene recordar que por más que algunos insistan, no todo el mundo festeja el Día de Acción de Gracias, ni Navidad, ni habla un mismo idioma, ni comprende lo mismo con un gesto. La Web, desde sus inicios, tiene una estructura internacional, por lo que en el transcurso del diseño de un sitio web nos encontramos con una de sus consecuencias, que es la forma en que debemos manejarnos ante las diversas culturas de los distintos usuarios.

Conviene recordar que, una vez que subimos nuestra primer página a Internet, y aunque hablemos en ella de temas meramente locales, será visitada por usuarios de lugares muy diversos, en donde la cultura, y todo lo que esto genera (modo de hablar, escribir, medir el tiempo, formas de comparación y ordenamiento, de usar el calendario, etc.) puede ser muy distinta.

¿Cómo hacer esto?

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

La internacionalización de un sitio web no se refiere solamente a crear un sitio multilingüe. Enumeremos los aspectos que debemos tener en cuenta:

1. Codificación de lenguas.
2. Negociación de contenidos.
3. Despliegue de caracteres e ideogramas.
4. Dirección del texto.
5. Uso semántico.
6. Ambigüedad lingüística.
7. Formatos de fechas y calendarios.
8. Preferencias del navegador.
9. Direcciones web multilingüales.
10. Comparación de cadenas o reglas de ordenación.

1. Codificación de lenguas

a. Normas

Cada lengua tiene (o debería tener) un código único que la identifique y diferencie del resto. Para esto, se intentó hacerlo a través de una norma de la Organización Internacional de Estándares (ISO). Esta norma es la ISO 639. Originalmente, fue una norma de 1988, que actualmente consta de 6 partes, algunas más consolidadas que otras. Además, hay otras normas que intentan complementar o mejorar al conjunto de ISO 639. La primera versión identificaba solamente 136 lenguas. La segunda 400 y ya comenzaba a identificar grupos de lenguas. Las últimas versiones intentan abarcar todas las lenguas existentes, además de los grupos y familias de lenguas.

b. Código

Una vez que sabemos de dónde sacar los códigos correspondientes a cada lengua, veamos cómo “hacer la etiqueta” para los distintos lenguajes de marcado:

HTML:

Si todo el texto de la página está escrito, por ejemplo, en guaraní, agregaremos en la etiqueta `<html>` lo siguiente:

```
<html lang="gn">
```

Donde *gn* es el código correspondiente al guaraní.

En el caso que insertemos en un texto de un idioma una palabra o frase de otro, deberemos agregar el atributo `lang` en la etiqueta *span*, por ejemplo:

```
<p>En mapuche, gente o pueblo se dice <span lang="arn">che</span>.</p>
```

XHTML:

Por compatibilidad entre HTML y XML, usaremos la siguiente sintaxis:

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

```
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="es"
lang="es">
```

XML:

Usaremos la etiqueta xml:lang de la siguiente forma:

```
<p>En mapuche, gente o pueblo se dice <span
xml:lang="arn">che</span>.</p>
```

2. Negociación de contenidos

a. Uso de HTTP

El protocolo HTTP es el más usado por la Web. De hecho, es el que, por defecto, nos permite ver y navegar las páginas web. Cuando un navegador se conecta con un servidor web remoto, no sólo le solicita la página cuya dirección (o URL) el usuario ha escrito (o la indicada por el enlace seleccionado) sino que le informa de detalles del navegador, de la página solicitada, etcétera. Este mensaje que envía el navegador al servidor web se llama "petición". Algunos de los parámetros que el navegador envía pueden ser modificados.

Luego, una vez que la conexión se realiza, el servidor también contesta enviando una respuesta estándar HTTP (indicando si la petición tuvo éxito o no) y a continuación la página solicitada. Esta respuesta del servidor también puede ser modificada si se tienen los permisos necesarios.

b. Uso de HTTP desde el servidor

Es importante asegurarse de que toda información sobre idioma y codificación de caracteres enviada por el servidor web sea correcta, dado que la información incluida en el encabezado HTTP anula la información que contiene el documento mismo. Es muy común que ocurran conflictos de codificación entre nuestro servidor web y las páginas cuando decidimos cambiarnos a UTF-8. Cuando un servidor web es instalado, el administrador le dice qué sistemas de codificación son aceptados y cuál es el sistema por defecto.

3. Despliegue de caracteres e ideogramas

a. Codificación de caracteres

Si utiliza un carácter diferente a la mayoría de los caracteres básicos necesarios para el idioma que habla, es posible que no se pueda leer su texto a menos que especifique la codificación de caracteres que utilizó.

La información de codificación inadecuada no sólo perjudica la capacidad de lectura de un texto que se visualiza, sino que, además, puede significar que sus datos no se encuentren en una búsqueda o que no se puedan procesar de manera confiable de diversas formas.

Las palabras y las oraciones de un texto se crean a partir de caracteres.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

Algunos ejemplos de caracteres incluyen la letra latina acentuada “á”, el ideograma chino “請” o el carácter devanagari “ॢ”.

Los caracteres se agrupan en un set o conjunto de caracteres. Luego, se denomina un set de caracteres codificados cuando a cada carácter se le asigna un número en particular, denominado punto de codificación. Estos puntos de codificación se representarán en la computadora por uno o más bytes.

El confuso término charset con frecuencia se utiliza para referirse a lo que realmente son las codificaciones de caracteres. Desafortunadamente, existen muchos sets de caracteres y codificaciones de caracteres diferentes, es decir, muchas maneras diferentes de mapear entre bytes, puntos de codificación y caracteres.

b. Despliegue de caracteres

Un parche (que ya no se recomienda): Los caracteres de escape:

- **á**; para á.
- **é**; para é.
- **ñ**; para ñ.

c. Declaración del conjunto de caracteres

El Consorcio W3 dice ¹ que todo documento transmitido por HTTP que sea de tipo texto (texto plano, HTML, etcétera) debe tener un parámetro charset que especifique la codificación de caracteres del documento.

```
<meta http-equiv="Content-type" content="text/html; charset=ISO-8859-1 " />  
<meta http-equiv="Content-type" content="text/html; charset=utf-8"  
>
```

(Ver el "Test del Ñandú ²" para mayor información sobre problemas de configuración de servidores web).

d. Unicode

Es un código de caracteres unificado. Casi todos los caracteres, glifos, ideogramas y símbolos que existen están incluidos en Unicode, y los pocos que quedan fuera están siendo incluidos con cada nueva versión.

La forma de codificación de datos de Unicode simplifica el desarrollo de aplicaciones y sitios web de uso internacional, porque permite que todos los caracteres internacionales estén representados con una única codificación.

4. Dirección del texto

1 <http://www.w3.org/International/O-HTTP-charset>.

2 En la página “ASCII, Unicode, UTF-8 y la Internacionalización - parte II” dentro del sitio Juque.cl <http://www.juque.cl/weblog/2006/04/02/ascii-unicode-utf8-y-la-iatarnaaianalizaaian-parte-ii.html>.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

Conviene recordar que existen lenguas que se leen de derecha a izquierda, como el hebreo, el árabe, el sirio, el pashto, el persa, el sindhi, el thaana, el urdu, el yiddish, etc.

Además, algunas lenguas ideográficas, como el japonés, el coreano, el mongol o el chino mandarín, son más flexibles en su dirección de escritura. En algunos casos se escriben de izquierda a derecha o verticalmente de arriba hacia abajo, aunque también se escriben (verticalmente) de derecha a izquierda. Algunos diarios chinos, por ejemplo, a veces combinan todas esas direcciones de escritura en una misma página³.



Ilustración 1 - Imagen de periódico chino en el que se ven diversos textos con diversas orientaciones.

¿Cómo trabajar con ellas?

a. Declarar Unicode para la página:

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

b. Si no todo el texto está escrito en una lengua, declarar el cambio de dirección cuando se produce en el código:

```
<p>Texto en español <span dir="rtl" lang="ar" xml:lang="ar"> مفتاح
```

```
معايير الويب</span> texto en español.</p>
```

c. También podemos agregar una traducción para el caso que algún usuario use un navegador que admita la opción **title**:

```
<p>Estos países son Barhein, <span dir="rtl" lang="ar" xml:lang="ar" title="Significa: 'Egipto'">مصر</span> y Kuwait.</p>
```

Además, conviene recordar que, como occidentales, la costumbre de ver una imagen o la estructura de un diario o sitio web de izquierda a derecha tiene que ver, en parte, con que aprendimos a leer en esa dirección. Por esto, también debemos pensar que deberemos cambiar la estructura de nuestro sitio web si el mismo está destinado a un público habituado a leer de derecha a izquierda.

3 La imagen que se ve es de una aportada por Richard Ishida en la siguiente dirección: <http://rishida.net/docs/unicode-tutorial/images/direction/bidi4.png>, que forma parte del artículo que está en la página <http://rishida.net/docs/unicode-tutorial/part4>.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

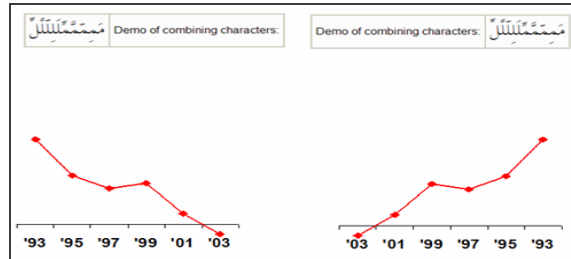


Ilustración 2- Un mismo gráfico es mostrado para un contexto occidental (izquierda) y para contenido árabe (derecha).

A esto último (inversión del diseño de las ventanas, los menús, los cuadros de diálogo y demás elementos visuales de un sitio web) se lo llama, en diseño web, **reflejo**.⁴



Ilustración 3 - Capturas de pantalla del sitio de la BBC, en sus versiones en español y árabe.

5. Uso semántico

El llamado "marcado semántico" tiene mucha importancia en diversas áreas del diseño web, como la accesibilidad y también en la internacionalización.

4 Un detalle que puede resultarnos curioso y que puede llegar a influir a la hora de diseñar una página, es que si justificamos un texto, no se mostrará de la misma forma en todas las formas de escritura. Mientras que en las lenguas occidentales, el justificado se realiza agregando espacio a los espacios entre palabras, en lenguas como el árabe, lo que se "estira" es la palabra.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

Así, por ejemplo, debemos recordar de tener en cuenta que si enfatizamos un texto, debemos usar la etiqueta y no la etiqueta <i>. Esto se volverá más evidente si el texto no está en una lengua occidental conocida, sino en otras como el japonés, donde los ideogramas etiquetados con <i> no mostrarán visualmente ninguna diferencia con el resto del texto. En cambio, si usamos , luego podremos darle con CSS2 un recuadro sombreado a esos ideogramas (opción *amikake*), o indicarle que muestre un pequeño círculo sobre cada ideograma con CSS3 (opción *wakiken*)⁵:

これは日本語です。

Ilustración 4 - Texto en japonés, sin énfasis.

これは日本語です。

Ilustración 5 - Mismo texto, con énfasis wakiken, a través de pequeños círculos sobre los ideogramas a resaltar.

これは日本語です。

Ilustración 6 - Mismo texto, con énfasis amikake, a través de un recuadro sombreado.

Ahora, si nuestro texto está en tibetano, pasará algo parecido. En esta lengua se usan dos pequeñas marcas debajo de cada sílaba enfatizada, por lo que un buen uso del marcado semántico nos ayudará⁶:

འཇུག་བཟུང་ནི་རྣམ་བུ་དྲི་མ་མེད་པའི་ཉེས།

Ilustración 7 - Texto en tibetano con sílabas enfatizadas.

En cirílico usado en el ruso, ucraniano y otras lenguas, el énfasis puede aparecer como un símbolo parecido (no igual) a la tilde de nuestra ñ (ñ, ŷ), el texto puede adoptar formas redondeadas, o puede cambiar completamente de forma.

6. Evitar la ambigüedad lingüística

Una palabra puede ser muy clara en un idioma, pero no tanto al traducirla a otro, por lo que debemos tener en cuenta esto a la hora de programar y diseñar.

5 El origen de las imágenes es <http://rishida.net/docs/unicode-tutorial/images/miscellaneous/misc.png>, que forma parte de la página <http://www.alistapart.com/articles/worldgrowssmall>, de Richard Ishida.

6 La imagen es provista por Richard Ishida, su dirección de origen es <http://people.w3.org/rishida/blog/images/tibetan-emphasis.gif> y forma parte de la página <http://rishida.net/blog/?p=101>.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

Por ejemplo, la frase “*Empty folder*” puede resultar un texto claro en inglés, pero puede resultar una construcción ambigua al intentar traducirla a otro idioma, ya que las cadenas se pueden traducir de forma diferente según las funciones gramaticales de los componentes de dichas cadenas. Así, “*empty*” puede ser tanto un verbo como un adjetivo y esto puede llevarnos a traducciones diferentes en idiomas como el italiano o el francés.

Se recomienda, dependiendo de cada caso, que los sitios web tengan referencias culturales e idiomáticas neutras.

7. Formatos de fechas y calendarios

El carácter internacional de Internet hace que nos encontremos con diversas formas de representar las fechas, además de los calendarios no gregorianos:

- Estados Unidos: 11/08/12 ó 11/08/2012
- Europa y parte de Latinoamérica: 08/11/12 ó 08/11/2012
- Japón: 12年11月8日 ó 2012年11月8日
- China: 2012年11月8日⁷

También está el caso de los que separan las cifras con guiones o puntos.

Esto nos recuerda que, por más que nos parezca común, nuestra forma de mostrar una fecha no es la común para todos en el mundo.

a. Calendarios

Actualmente coexisten en el mundo más de 40 calendarios⁸. Algunos de ellos son el gregoriano, el hebreo, el árabe, el japonés, el taiwanés, el coreano, el chino, el tailandés, el budista, el persa, etc. ¿Cómo atender a las necesidades de tal variedad de usuarios?

Para tener en cuenta tal variedad, se recomiendan tres posibilidades:

- I. Usar un formato local neutro.
- II. Explicitar los meses y años hasta la obviedad.
- III. Usar la cabecera de HTTP Accept-Language.

I. Usar un formato local neutro

La norma ISO 8601 especifica que las fechas se deben mostrar en el formato AAAA-MM-DD. Esto tiene cosas a favor y en contra.

A favor:

- No resulta ambiguo.
- Es amigable para los usos informáticos que deseemos darle a la fecha (bases de datos, búsquedas, ordenamientos, web semántica, etc.).

⁷ La fechas pueden también aparecer como el formato europeo, pero no es lo más común (Mazel Sussman, Julie, “*I can read that! A traveler's Introduction th Chinese Characters*”, página 31).

⁸ Un detalle de los mismos se puede leer en la página http://es.wikipedia.org/wiki/Calendario#Tipos_de_calendario.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

En contra:

- No es un formato amigable para las personas.
- Ocupa más espacio que una fecha en formato común.
- Resulta confuso para los usuarios de países que tienen un calendario alternativo, como por ejemplo en Tailandia, donde usan el calendario budista.

II. Explicitar los meses y años hasta la obviedad

Usar nombres para los meses (abreviados o no) y 4 dígitos para los años. Esto también tiene cosas a favor y en contra.

A favor:

- No resulta ambiguo.
- Es amigable para los usuarios.

En contra:

- Es muy poco amigable para los usos informáticos.
- Ocupa más espacio.
- En algunas lenguas el uso de tres letras para el mes puede resultar problemático (por ejemplo, en francés, Junio y Julio son Juin y Juillet).
- No siempre los meses serán fácilmente reconocibles en distintas lenguas (para un español no hablante de inglés, Jan no le significará Enero).

III. Usar la cabecera de HTTP Accept-Language

La cabecera HTTP **Accept-Language**, tal como comentamos antes, especifica las preferencias de la lengua del usuario, pero también es usada para determinar preferencias locales. Esto también tiene cosas a favor y en contra.

A favor:

- No resulta ambiguo.
- Es amigable para los usuarios.

En contra:

- Este método funciona principalmente para páginas web dinámicas, donde se inserta la fecha desde un almacenamiento externo.
- No todos los servidores web permiten configurar esta opción.

8. Preferencias del navegador

Cuando un navegador pide un documento a un servidor web, le informa sobre las preferencias de idioma a través de la cabecera HTTP **Accept-Language**. Si el servidor tiene varias versiones de una página en distintos idiomas, esta información puede usarse para recuperar la página en el idioma preferido, siempre que exista. Si hay una única versión de dicha página en ese servidor, esa es la que se recuperará.

La mayoría de los navegadores permiten cambiar las preferencias del idioma, por ejemplo:

es para español y **es-AR** para el español de Argentina.

Si tenemos un navegador en nuestro idioma, la mayoría de las veces se instala con la configuración inicial deseada.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

En algunos navegadores es posible establecer más de un idioma, dándoles un orden de preferencia, como se muestra a continuación:

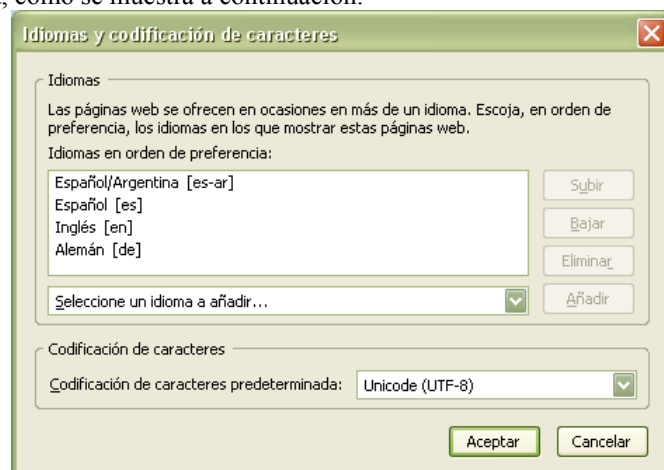


Ilustración 8 - Opción de idiomas y codificación de caracteres en Mozilla Firefox.

Una vez realizados los cambios, se puede probar su funcionamiento (o no) con sitios que tienen negociación dinámica de idioma, como Google o Debian.

9. Direcciones web multilingüales

Una dirección web es usada para apuntar a un recurso en la Web, como una página web. Estas direcciones se expresan a través de Identificadores Uniformes de Recursos o URIs. La sintaxis de las URIs se define en la norma RFC 3986 STD 66, la cual restringe las direcciones web a un pequeño número de caracteres:

- Letras mayúsculas y minúsculas del alfabeto en inglés.
- Numerales europeos.
- Un pequeño número de símbolos.

Esto se debe a que el ámbito en el que la Web comenzó a desarrollarse era el angloparlante. Pero al poco tiempo se vio que este ámbito se ampliaba a pasos agigantados haciendo necesario el tener que usar caracteres de cualquier lenguaje en la direcciones web. Una dirección web en nuestra propia lengua es fácil de crear, recordar, transcribir, interpretar y relacionar. Y esto nos lleva a mejores negocios, mejorar las búsquedas y las comunicaciones. En resumen, mejora la Web.

Para lograr esto existe algo llamado Identificador de Recursos Internacionalizado o IRI.

Actualmente varios formatos soportan IRI, como HTML 4.0 y 5, identificadores de sistema de XML 1.0, el atributo href de XLink, cualquier tipo de datos de URIs del Schema de XML, etc.. Desafortunadamente, no muchos protocolos soportan IRI. Así, es posible encontrar una dirección como la que sigue:

http://納豆.例.jp

En donde la declaración del protocolo (**http://**) debe estar en caracteres ASCII, mientras que el dominio es una mezcla de caracteres ASCII con japoneses.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

Pero esto no sólo es necesario para lenguas orientales, como se puede ver en la siguiente dirección sueca:

<http://räksmörgås.josefsson.org/mål/franzén.html>

10. Comparación de cadenas o reglas de ordenación

Supongamos que desde nuestro sitio ofrecemos la opción de mostrar, de acuerdo a las preferencias del usuario, datos ordenados con diversos parámetros.

Hasta aquí, para cualquier persona que sepa algo de manejo de bases de datos, no habría problemas... ¿no?

Bueno, no habría problemas si nos guiáramos solamente con nuestra lengua y nuestro calendario. Pero si queremos tener en cuenta otras culturas, nos encontraremos con algunas cositas más a tener en cuenta.

a. Fechas y calendarios

De acuerdo a lo que leímos anteriormente, no hay UNA fecha ni una sola forma de mostrarla, así que esa es una de las cuestiones básicas a tener en cuenta. Si vamos a elegir una de las formas y calendarios para mostrarlos ordenadamente, y siguiendo las pautas de accesibilidad a la Web, deberemos informar de este parámetro al usuario que hace la consulta y, dentro de lo posible, darle la posibilidad de elegir otros calendarios o formatos.

b. Caracteres distintos a los que usamos

De acuerdo a nuestra experiencia cotidiana, ya sabemos que si nos piden ordenar determinados nombres en orden alfabético, deberemos poner la A antes que las B (salvo que nos pidan en orden inverso, claro) y las únicas posibles dudas que surgirían serían cuando no sepamos si considerar a la CH o la LL como letras únicas o no.

Pero... ¿qué pasa si nuestros posibles clientes o usuarios hablan chino mandarín, japonés o hindi, por citar tres lenguas MUY populares?

Bueno, deberemos tomarnos el trabajo de averiguar qué sistema usa cada cultura.

El chino mandarín, por ejemplo, tiene dos formas distintas de ordenar (obviamente que ninguna es “alfabética”, ya que no tiene un alfabeto, tal como lo entendemos los occidentales): una por pronunciación y otra por el número de trazos. La elección de uno u otro depende de la región del usuario, porque los usuarios de Taiwán y Hong Kong preferirán la ordenación por trazos, mientras que los usuarios de China Continental, Singapur y Macao se inclinarán por la ordenación por pronunciación.

Y tampoco debemos irnos muy lejos en las distancias, para igualmente tener problemas de ordenamiento.

Mire su teclado y vea algunos caracteres que aparecen: ß, Æ, â, ë, ñ, ç y muchos otros, dependiendo del origen del teclado o el marcado al que va dirigido.

Algunos, como la ñ o la ç, sabemos, o imaginamos en dónde deberemos colocarlos a la hora de ordenar.

Pero... ¿estamos seguros de dónde irán ß o Æ?

Tomemos la ß. No es la letra griega beta (β) ni una B deformada. Se llama ezeta, ese-zeta o ezsett. Es de uso muy común en los idiomas alemán y danés. Hasta aquí, parece algo relativamente normal, pero la cosa comienza a complicarse cuando nos

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

enteramos que es una letra minúscula y no hay versión en mayúscula. Se usa la misma letra, con la misma forma, tanto en un texto en mayúscula como en minúscula. Esto es algo que deberemos tener en cuenta si una decisión de seguridad, por ejemplo, se basa en el resultado de una comparación de cadenas o una operación de cambio de mayúsculas y minúsculas.

Además, en alemán la ß y la ss son consideradas equivalentes.

Ahora veamos el carácter Æ. Este carácter (que sí tiene diferencias entre mayúsculas y minúsculas: Æ y æ) se usa en los idiomas inglés, alemán, sueco y danés. Pero mientras que en el alemán se lo ubica luego de la A, en el danés y sueco, se lo ubica luego de la Z. Y otra diferencia surge en la forma en que se lo trata. En danés, se lo trata como carácter individual, pero en inglés como símbolo especial, así que las palabras “Apple” y “Æble”, en danés irán en ese mismo orden, mientras que en inglés será al revés, ya que a los símbolos especiales, como el que nos ocupa, se los ordena antes de la A en el alfabeto.

El uso de las diéresis también tienen sus características. En el sueco, crean caracteres que son vistos como letras independientes del alfabeto, que aparecen luego de la Z, mientras que en francés su uso es distinto.

Otro ejemplo es el uso de la secuencia “ch” en checo, ya que es usada como una unidad simple y única para ordenar, así que las palabras comenzadas con “ch” van luego de la “h” y no luego de la “c”.

¿Eso es todo?

No, es sólo el comienzo.

Falta todavía recordar que las diferencias culturales pueden a veces resultar más sutiles o a veces pasar un poco desapercibidas.

a. Colores

Por ejemplo, conviene recordar que el uso de los colores puede resultar distinto. Así, si vamos a tocar temas relacionados con el luto, conviene recordar que para algunas culturas, como la japonesa, el blanco representa el luto, y no el negro. De la misma forma que no es lo común que una novia se vista de negro en Occidente, cosa que no es raro en Oriente ⁹.

b. Símbolos

No siempre los símbolos tienen el mismo significado en todo el mundo.

Un buen ejemplo es nuestra conocida marca de tilde (√). En algunos países significa *incorrecto*, no lo opuesto ¹⁰.

9 Ishida, Richard, “Color”, en presentación “*Introduction to Writing Systems*”, diapositiva 39, 10 de junio de 2003 <http://www.w3.org/2006/Talks/04-i18n-china.pdf>.

10 Ishida, Richard, “Color”, en presentación “*Introduction to Writing Systems*”, diapositiva 75, 10 de junio de 2003 <http://www.w3.org/2006/Talks/04-i18n-china.pdf>.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

Otra diferencia a tener en cuenta es que en Japón, los círculos (○) reemplazan estas marcas de tilde, tal como se ve en la siguiente imagen ¹¹:



Ilustración 9 - Fotografía de un comercio japonés, en donde se ve el uso de círculos, en vez de tildes.

c. Gestos

También el lenguaje corporal o los gestos pueden dar lugar a problemas. Cada uno de los siguientes gestos pueden resultar ofensivos en alguna parte del mundo ¹²:



d. Direcciones postales

¿Estamos seguros que todos ponen los datos en una carta de la misma forma? ¿Cómo manejaremos los datos cuando el usuario no es nuestro vecino occidental?

Para un ruso o un japonés, las direcciones en las cartas se escriben de lo general a lo específico, lo opuesto a nosotros. Así que ellos pondrán primero el país, luego la provincia, el código postal, la localidad, la calle y número y, por último, el nombre del destinatario. Recordemos esto cuando ofrezcamos, a través de una base de datos, formularios de ingreso de datos, etiquetas postales u otra forma relacionada.

Otro tema es el uso del nombre y el apellido. Los chinos, y otros pueblos, cuando escriben su nombre, ponen primero su apellido y luego su (o sus) nombre(s). Por

11 Ishida, Richard, “Symbolism, color, graphics...”, en presentación “Introduction to Writing Systems”, diapositiva 76, 10 de junio de 2003
<http://www.w3.org/2006/Talks/04-i18n-china.pdf>

12 Ishida, Richard, “Symbolism, color, graphics...”, en presentación “Introduction to Writing Systems”, diapositiva 78, 10 de junio de 2003
<http://www.w3.org/2006/Talks/04-i18n-china.pdf>

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

ejemplo Mao Zedong (o Mao Tsé Tung, para el viejo sistema de traducción), tenía por apellido Mao, no Zedong, lo mismo que Deng Xiaoping (apellido Deng). Pero si esos ciudadanos chinos se mudaban a otro país donde la costumbre fuera poner primero el nombre, ellos mismo, muchas veces, lo hacían también, para evitar confusiones, por lo que podríamos cometer la distopía de imaginarnos un Mao que se mudara a Occidente y se podría haber hecho llamar Zedong Mao¹³.

e. Localismos

Otro tema es el localismo, así que si en nuestros textos, información o publicidades vamos a hacer referencia a chistes que se ven en la televisión o se oyen en la radio de nuestra ciudad, es muy probable que el otro se “quede afuera” y no entienda (o peor: entienda mal y se ofenda) el sentido de lo que queremos transmitir.

En resumen

Se podría resumir todo con la frase “tener en cuenta al otro”.

Esto lo podemos tomar de dos formas. Por un lado, como “otro problema” en el que tendremos que adentrarnos para intentar solucionar las necesidades de los usuarios en aumento de nuestro sitio web.

Pero por otro lado, es una oportunidad de ver el problema como el comienzo de nuevas y apasionantes experiencias, en donde, a medida que nos comencemos a interiorizar por la cultura “del otro”, también comenzaremos un viaje interior, a nuestra propia cultura, y si tenemos la suficiente paciencia, quizás descubramos (como a mí me ha pasado) que en el fondo no hay un “ellos” y “nosotros”, sino que vivimos en una sociedad mundial compleja y variadísima... pero única, fascinante y siempre dispuesta a enseñarnos algo nuevo y alternativo con qué enriquecer nuestra pequeña existencia.

Bibliografía:

Libros:

Mazel Sussman, Julie, "*I can read that! A traveler's Introduction th Chinese Characters*", China Books & Periodicals, Inc., 1994, San Francisco, Estados Unidos, ISBN 0-8351-2533-5.

Ortiz, Renato, "La supremacía del inglés en las ciencias sociales". Siglo Veintiuno Editores, 2009, ISBN 978-987-629-080-7.

Ouane, Adama (editor), "*Towards a Multilingual Culture of Education*", UNESCO, 2003, ISBN 92 820 1131-3.

Xi, Xiandai, Han, Han, Cidian, Xi, "Diccionario moderno español - chino, chino - español". 1992. ISBN 7-5600-0779-1/H.335.

Documentos electrónicos:

Crystal, David, "*Scope of Internet linguistics, The*", 2005.

Holzschlag, Molly E., "*Internationalization: awakening the sleeping giant*", 2006.

Ishida, Richard, "*Practical & Cultural Issues in Web Design*", 2003.

13 Mazel Sussman, Julie, "I can read that! A traveler's Introduction th Chinese Characters", página 67.

Internacionalización web: Un conjunto de herramientas para la democratización de la información y el conocimiento.

Ishida, Richard, "*Internationalizing the Web*", 2005.

Mufwene, Salikoko S., "*Globalization and the Myth of Killer Languages: What's Really Going on?*".

Ruiz Vieytez, Eduardo J., "Lenguas oficiales y lenguas minoritarias: cuestiones sobre su estatuto jurídico a través del derecho comparado", 2004.

UNESCO, "Plurilingüismo", 2006.

UNESCO, "Recomendación sobre la promoción y el uso del plurilingüismo y el acceso al ciberespacio".

Valero, Pedro, "Internacionalización", Universitat de València.

Artículos en revistas y diarios:

Crystal, David, "*Weaving a Web of linguistic diversity*", The Guardian, 2006.

Sitios web:

Ethnologue (<http://www.ethnologue.com>), dependiente de la Organización SIL International (<http://www.sil.org>). Gordon, Raymond G., Jr. (editor), 2005.

Decimoquinta edición. Dallas, Texas.

Globalización y localización (<http://msdn.microsoft.com/es-es/library/ms753931.aspx>).

Internationalization, W3C (<http://www.w3.org/International>).

Language Tags (<http://www.lagtag.net>).

Richard Ishida (<http://rishida.net>).

Wikipedia en español (<http://es.wikipedia.org>). También hay versiones de Wikipedia en 284 lenguas más. Un listado de las mismas se puede encontrar en:

<http://es.wikipedia.org/wiki/Anexo:Wikipedias> .

Artículos o páginas en sitios web:

Arko, Phil, "*International & multilingual web sites*"

(<http://www.w3.org/International/questions/qa-international-multilingual>).

BCP 47 links (<http://www.w3.org/International/core/langtags/rfc3066bis>).

Bos, Bert, "Codificación de caracteres" (<http://www.w3.org/International/O-charset>).

Dürst, Martin, "Codificación de formularios plurilingües"

(<http://www.w3.org/International/questions/qa-forms-utf-8>).

"Guía breve de internacionalización"

(<http://www.w3c.es/Divulgacion/GuiasBreves/internacionalizacion>).

Holzschlag, Molly E., "*World Grows Small: Open Standards for the Global Web*",

2006 (<http://www.alistapart.com/articles/worldgrowsmall>).

Honomichl, Lloyd, "Formatos de fechas"

(<http://www.w3.org/International/questions/qa-date-format>).

Korpela, Jukka "Yucca", "*A tutorial on character code issues*",

(<http://www.cs.tut.fi/~jkorpela/chars.html>).

Phillips, Addison, "*Understanding the New Language Tags*"

(<http://www.w3.org/International/articles/bcp47/Overview>).

Ruby, Sam, "*intertwingly*" (<http://intertwingly.net/stories/2004/04/14/i18n.html>).

Spolsky, Joel, "*The Absolute Minimum Every Software Developer Absolutely,*

Positively Must Know About Unicode and Character Sets (No Excuses!)"

(<http://www.joelonsoftware.com/articles/Unicode.html>).

Yergeau, François, "*When to use language negotiation*"

(<http://www.w3.org/International/questions/qa-when-lang-neg>).