

Localización e internacionalización de sitios web

Noelia Corte

MSc en Electronic Publishing (City University, Londres)

Resumen

La única manera de utilizar la Internet de forma rápida y eficaz para llegar a los mercados extranjeros es mediante un correcto proceso de localización. Aunque ésta es una actividad relativamente nueva y la mayoría de las herramientas y procedimientos a seguir aún están desarrollándose, existe una serie de elementos comunes que serán analizados en este artículo.

Palabras clave

localización, internacionalización, sitios web, Internet.

Introducción

Desde sus inicios, el inglés ha sido el idioma predominante de la *World Wide Web* ya que ésta era dominio exclusivo de los mercados estadounidenses y británicos. Sin embargo, en los últimos años se ha producido un aumento en el número de usuarios cuyo primer idioma no es el inglés. Aunque el inglés se acepte como la *lingua franca* de la Internet y la mayoría de sus usuarios lo entiendan, es natural que un sitio web en su idioma materno les resulte mucho más atractivo y fácil de utilizar.

Los potenciales clientes que navegan por Internet a diario pasan más tiempo y están más dispuestos a comprar productos en aquellos sitios que están en su idioma materno. Así que toda empresa deberá en primer lugar salvar las barreras del idioma si desea sacar el mayor provecho del inmenso potencial que la Internet ofrece para acceder a diferentes mercados.

1. Terminología

Tres términos son claves a la hora de entender en qué consiste el proceso de localización de un sitio web:

1.1 Internacionalización

Consiste en la identificación de toda la información local que aparece en un sitio web, es decir, aquella información que viene dictada por el idioma y la cultura del país donde se diseñó originalmente. Por ejemplo fechas, números, moneda, información de contacto, etc. Estos elementos deberán aislarse y guardarse de forma independiente para que sea posible adaptarlos a las especificaciones de cualquier idioma.

1.2. Localización

Es el proceso de adaptar un sitio web a un idioma y una cultura diferente. Esto significa mucho más que simplemente traducir el contenido de las páginas. El contenido de una página web está formado por texto e imágenes, ambos deben ser traducidos y sometidos a una adaptación cultural. El usuario nunca debe notar que ese sitio fue originalmente creado en otro idioma.

1.3. Globalización

La globalización combina los procesos de internacionalización y localización. Consiste en el diseño de sitios web que pueden ser utilizados en diferentes países con un mínimo de cambios. Es un concepto que pertenece más al área del marketing que al área técnica.

La internacionalización y localización de sitios web son procesos complicados y costosos, por lo que muchas empresas son reacias a llevarlos a cabo. Sin embargo, los beneficios a largo plazo sobrepasan con creces los costes iniciales: creación de nuevos mercados, aumento de ventas, simplificación del proceso de actualización de la información, mayor rapidez a la hora de introducirse en nuevos mercados, etc.

A continuación se analizarán con carácter general los diferentes pasos a seguir para internacionalizar y localizar con éxito un sitio web.

2. Análisis del Mercado término y del sitio web original

Antes de comenzar el proceso de localización es esencial conocer a fondo el mercado término.

- Quiénes son los usuarios;
- Qué esperan obtener del sitio;
- Qué espera obtener la empresa del sitio localizado (¿Simplemente ofrecer más información? ¿Crear una presencia en mercados extranjeros con vistas a abrir nuevas oficinas? ¿Aumentar sus ventas? Si es así, ¿va a ofrecer la posibilidad de comprar artículos en línea?);
- Quiénes son sus competidores en el mercado término;
- Requisitos legales específicos de un país determinado;
- Características culturales, religiosas o sociales que puedan afectar la presentación del contenido (a veces puede resultar beneficioso contratar los servicios de un antropólogo si la cultura es muy diferente a la original);
- Requisitos técnicos, bando de ancha disponible, precios de conexión a Internet;
- Requisitos para registrar nombres de dominio;
- Legislación en materia de protección de la información personal y copyright.

Los departamentos legal y de marketing deberían colaborar estrechamente con los expertos del país término y preparar un documento que cubra todas estas áreas para cada uno de los idiomas a los que se vaya a localizar el sitio web. El jefe de proyecto de localización debe tener un conocimiento profundo de lo que es lo que se espera del sitio web ya que estos requisitos iniciales determinarán la duración y el coste del proyecto.

En el mundo del comercio electrónico no es suficiente con llevar a cabo un Análisis de Mercado en el sentido tradicional. Además se deberá investigar cuántas personas poseen ordenadores y el nivel de uso de Internet en el mercado término, cuántos usuarios poseen banda ancha, los precios de conexión en los diferentes países, si los usuarios tienen acceso a las últimas tecnologías, niveles de ventas de hardware y software, etc.

Con los objetivos bien establecidos y un detallado conocimiento de los mercados en los que se desea trabajar, llega el momento de analizar el sitio a localizar, tanto su contenido como su funcionalidad. En la actualidad la mayoría de los sitios web se diseñan con vistas a ser utilizados en un solo idioma por lo que plantean muchos más retos para su localización.

Antes de comenzar la localización el sitio deberá ser internacionalizado para facilitar posibles futuras localizaciones a otros idiomas.

3. Internacionalización

Todos los elementos locales del sitio deberán ser identificados y aislados. El proceso de internacionalización en muchos casos requiere drásticos cambios en el diseño y la funcionalidad del sitio. Esto puede ser costoso y llevar mucho tiempo pero una vez el sitio está internacionalizado, podrá localizarse a cualquier otro idioma con gran facilidad. El proceso de internacionalización sólo debe llevarse a cabo una vez y a largo plazo supondrá un gran ahorro en los planes de expansión al extranjero de cualquier empresa.

Otras consideraciones a tener en cuenta en este punto son:

- Cuáles son los idiomas a utilizar y a qué países queremos llegar. Es más fácil trabajar con idiomas europeos occidentales, por ejemplo, que con el ruso que utiliza un alfabeto distinto o el árabe que se escribe de derecha a izquierda.

- Cómo se llevarán a cabo las actualizaciones de contenido. Un sitio web que no actualiza sus contenidos nunca tendrá éxito, los usuarios de Internet buscan información rápida y actual. El mejor modo de controlar las actualizaciones de contenido y la gestión diaria del sitio es mediante la utilización de un sistema de *workflow* y gestión de contenidos.

Los gráficos deberán ser analizados y todos aquellos que tengan texto superimpuesto deberán, o bien ser eliminados, o las versiones originales en capas deberán almacenarse en un directorio central para poder traducir la capa de texto.

Todo símbolo, color e imagen deberá contrastarse en la cultura término. Los iconos deberán ser tan neutros y genéricos como sea posible para poder ser reutilizados.

En este punto los archivos de la interfaz de usuario, contenido y código deberán separarse en diferentes archivos de recurso. Se pueden utilizar *Server Side Includes* (SSIs) para facilitar la tarea de actualizar el contenido sin afectar el diseño gráfico de la página.

Una vez se hayan creado los archivos de recurso es fácil calcular la magnitud del proyecto, el número de archivos, número de palabras (esencial para estimar los costes y duración de las traducciones), etc.

La funcionalidad del sitio web (que será común en todos los idiomas) y la interfaz de usuario deberán separarse claramente. Es importante diferenciar entre lo que se debe traducir (lo que ve el usuario) y la funcionalidad del sitio, que normalmente permanece oculta y, por lo tanto, no necesita ser localizada. Todo contenido localizable se extraerá del código fuente y se guardará en archivos de recurso como *dynamic link libraries* (DLL) con un solo archivo ejecutable. Este es sólo un ejemplo de la forma que este sistema puede adoptar:

```
app.exe
  lng_res.dll
  ale_res.dll
  fra_res.dll
  esp_res.dll
  etc.
```

En esta etapa del proyecto es muy importante que los programadores trabajen conjuntamente con los traductores y que añadan tantas explicaciones como sea posible en el código fuente. Del mismo modo, los programadores no deberán nunca hacer suposiciones sobre cómo funcionan las reglas gramaticales de un idioma o sobre cómo interpreta un idioma la información.

La estructura del sitio deberá planificarse antes de comenzar el proceso de localización. Uno de los mayores problemas es decidir cómo se van a combinar las diferentes versiones. El método más común consiste en guardar cada idioma en el mismo servidor puesto que así se pueden sincronizar las actualizaciones en todos los idiomas a la vez. Normalmente este tipo de sitio tendrá un enlace a cada uno de los idiomas en su página inicial con lo que sólo se necesita una URL que se puede utilizar de forma global.

Otra posibilidad es guardar cada sitio localizado en su propio país término. Este proceso hace más lento el mantenimiento pero disminuye los tiempos de descarga. Una tercera opción consiste en crear sitios diferentes para diferentes países. Esto ocurre cuando los nombres de dominio son totalmente diferentes si se traducen, el contenido es totalmente dependiente de una cultura determinada o cuando se utilizan diferentes diseños.

El objetivo final es crear un sitio con una infraestructura global que será gestionada centralmente por un único equipo técnico con módulos localizables que deberán ser traducidos y adaptados a los diferentes idiomas.

4. Localización

El primer paso en el proceso de localización es la identificación de los diferentes elementos que deben ser localizados:

- La interfaz de usuario
- La funcionalidad técnica
- El contenido

Estas tres áreas están íntimamente relacionadas y es esencial prestar detenida atención a cada una de ellas si se quiere conseguir una completa y satisfactoria localización. De nada sirve un sitio web donde el texto está perfectamente traducido si la funcionalidad del sitio está comprometida.

4.1. Interfaz de Usuario

Las Hojas de Estilo (CSS) se pueden utilizar de forma muy eficaz para cambiar la imagen visual de una página. Una de sus mejores características es que permiten cambiar el tipo de letra de un sitio web con un solo cambio en un único documento. Es recomendable escribir una hoja de estilo para cada uno de los idiomas que no tenga alfabeto latino y averiguar si esos idiomas cuentan con tipos de letra similares a los utilizados en el sitio original.

Los gráficos y tablas deberán tener bastante espacio libre para acomodar con facilidad las diferentes longitudes de los diferentes idiomas.

Otro aspecto a tener en cuenta es la cantidad de gráficos que vamos a incluir en nuestra página. Una fantástica página con animaciones, fotografías y complicados gráficos puede descargar en unos segundos a través de una conexión de banda ancha pero puede tardar varios minutos o incluso horas a través de un módem de 56k, o resultar totalmente imposible en un país con malas conexiones telefónicas.

4.2. Funcionalidad Técnica

Aunque la localización de software es una actividad habitual en la industria de la Tecnología de la Información, la localización de sitios web es mucho más reciente y aún está en proceso de desarrollo. Sin embargo muchos de los conceptos utilizados en la localización de software son también de aplicación en los sitios web.

Los sitios web actuales son estructuras multi dimensionales con contenidos tanto estáticos como dinámicos soportados por bases de datos y que utilizan una combinación de varios lenguajes de programación. Esta complejidad no hace más que aumentar gracias a la producción de ordenadores más potentes y las posibilidades que ofrece la banda ancha.

Los problemas técnicos vienen en gran medida dictados por los idiomas en los que se vaya a localizar el sitio.

4.2.1. Representación de caracteres.

Uno de los principales problemas técnicos que plantea la localización de sitios web es la representación de caracteres. El básico *American Standard Code for Information Interchange (ASCII)* cubre tan solo 128 caracteres, los justos para representar todos los caracteres del idioma inglés pero claramente insuficientes para el resto de los idiomas.

En la actualidad Unicode ofrece una cobertura mucho mayor y más sofisticada que es independiente del Sistema Operativo utilizado.

HTML utiliza un sistema más completo denominado Universal Character Set (UCS). Ha sido definido por la ISO10646 y es equivalente a Unicode. Ambas normas son actualizadas regularmente para añadir nuevos caracteres.

Los *Transformation Formats* (UTFs) se crearon para subsanar los problemas de compabilidad que presentaba el UCS con muchas aplicaciones y protocolos. Existen diferentes versiones, la más utilizada en la web es el UTF-8 que es compatible con ASCII siempre que no haya más de 127 caracteres. Un documento de HTML en UTF-8 continua siendo un documento ASCII/ISO-8859-1 (inglés, idiomas europeos occidentales, suahili y hawaiano) normal. Si contiene caracteres Unicode por encima de la posición 255 deben indicarse como unidades "&", por ejemplo "é" se representaría como é.

El protocolo "http" también ofrece el parámetro "charset" que identifica un método de convertir una secuencia de bytes en una secuencia de caracteres. El parámetro "charset" ajusta el navegador del usuario para que muestre los caracteres correctos. La mayoría de los navegadores seleccionan ISO-8859-1 por defecto, por eso es obligatorio incluir el parámetro correcto para todos los idiomas que no cubre el ISO-8859-1.

Idealmente se debería utilizar Unicode o UTF-8 pero esto conllevaría en algunos casos considerable re-programación del servidor y no todos los navegadores soportan UTF-8.

4.2.2. Idiomas con requisitos especiales

- Idiomas asiáticos

Aunque en la mayoría de los idiomas un único carácter puede expresarse con un solo byte (1-256), existen tantos caracteres únicos en chino o japonés que el sistema necesita 2 bytes (1-65.536) para encontrar uno. Estos caracteres se guardan en varias tablas, el primer byte indica al sistema operativo en qué tabla buscar y el segundo indica el lugar que ese carácter ocupa en la tabla. Esto constituye el *Double Byte Character Set (DBCS)*. Si se va a localizar al chino, japonés o coreano por ejemplo, el servidor deberá ser reprogramado para reconocer el DBCS.

Otro aspecto a tener en cuenta con estos idiomas es que no tienen espacios entre los distintos caracteres. Esto puede causar problemas con, por ejemplo, la utilización de saltos de línea.

- Idiomas que se escriben de derecha a izquierda

La etiqueta DIR (Algoritmo Direccional Unicode) indica la dirección en que se debe leer un texto. Esta etiqueta es necesaria para idiomas como el árabe, el hebreo o el urdu.

A veces es necesario combinar textos de derecha a izquierda con textos de izquierda a derecha. Cuando esto ocurre es necesario utilizar el algoritmo bidireccional (BIDI).

4.2.3. HTML

Todo el contenido localizable de una página deberá ser enviado a un traductor. La manera más segura de evitar que por error un traductor altere la funcionalidad de esa página es bloquear todos aquellos elementos que no tengan que traducirse. Esto es posible en la mayoría de los editores de HTML y de las herramientas de traducción asistida por ordenador.

Un aspecto a tener en cuenta es que el sistema operativo que se utilice al llevar a cabo la localización afectará algunos de los comandos de HTML. Por ejemplo, los botones de "Enviar" y "Borrar" en un formulario aparecerán como "Submit" y "Clear" en un ordenador con Sistema Operativo inglés.

4.2.4. Orden alfabético

Cada idioma utiliza un algoritmo de secuencia diferente para ordenar la información. Las listas alfabéticas serán diferentes según el idioma. La mejor solución es que los programadores cambien el orden alfabético como parte de la localización de la funcionalidad.

4.3. Contenido

Este es el elemento clave de cualquier sitio web. La principal funcionalidad de la web es la información y por lo tanto éste es el elemento principal de cualquier proyecto de localización.

4.3.1. Texto

Es vital que el texto sea traducido por traductores nativos del idioma al que se va a localizar el sitio y que sean especialistas en el tema a tratar para utilizar la terminología correcta. Además de ser lingüistas, también deben tener un conocimiento detallado de la cultura del país.

Los sitios web proporcionan dos tipos de contenidos:

- Estático: Es permanente y no necesita ser actualizado con regularidad, normalmente aparece en páginas web en HTML, ASP, etc. con funcionalidad limitada. La localización puede llevarse a cabo con un editor de HTML o una herramienta de traducción.

- Dinámico: El mejor modo de guardar este tipo de texto es en una base de datos que facilite su mantenimiento y actualización. El proceso ideal de localización se llevará a cabo con un sistema que identifique automáticamente el contenido actualizado y lo envíe a través de un sistema de *workflow* predefinido.

Como regla general es aconsejable evitar coloquialismos, regionalismos, juegos de palabras o chistes que suelen resultar muy difíciles de reproducir en otro idioma o cultura.

Una vez haya finalizado el proceso de traducción, será necesario realizar una corrección de pruebas en todas las páginas para asegurar la mejor calidad posible. Este control de calidad deberá mantenerse con controles regulares cada cierto tiempo e identificación de las secciones que vayan a ser actualizadas frecuentemente.

4.3.2. Imágenes gráficas

Toda imagen gráfica debe ser contrastada culturalmente. Todo símbolo que sea dependiente de una cultura específica debe evitarse.

Como regla general sería recomendable evitar todo símbolo religioso o festivo, partes del cuerpo o gestos con la mano, imágenes que simbolizan un significado textual (por ejemplo utilizar la imagen de una tabla como icono que simbolice una tabla gráfica), símbolos nacionales, mapas con fronteras polémicas. Se debe tener especial cuidado con el uso de los colores (por ejemplo, el negro, considerado en occidente como un color neutro y elegante simboliza la muerte en China) o imágenes de personas y animales (totalmente inaceptables en los países musulmanes, por ejemplo).

4.3.3. Aspectos legales

La legislación sobre copyright o protección de datos personales, nombres de dominios, condiciones de uso, etc. varía de unos países a otros. Es vital contar con el asesoramiento adecuado si se quieren evitar problemas legales.

Una vez se haya localizado la interfaz de usuario, el contenido y el código habrá que integrarlo todo en el sitio web localizado. El modo en que se realice dependerá del modo en el que se haya extraído el contenido y el código fuente. Lo más importante es que la estructura del sitio facilite una sencilla integración del contenido y que existan procesos adecuados para comprobar la integración antes de la publicación del sitio.

5. Control de calidad

Una vez el sitio está localizado éste debe pasar un riguroso control de calidad. Si a los usuarios no les gusta lo que ven en la pantalla sólo les llevará unos segundos salir del sitio y

no volverán nunca más. Algunos de los peores problemas que aparecen en sitios que supuestamente ya han sido localizados son:

- Texto en el idioma original, sobre todo en mensajes de error o cuadros de diálogo;
- faltas de ortografía;
- errores gramaticales;
- terminología errónea (sobre todo en sitios web técnicos o especializados);
- imágenes incongruentes.

La mejor solución para evitar estos problemas y mantener los costes al mínimo es realizar una versión general, una vez su funcionalidad ha sido comprobada el texto puede adaptarse a los mercados particulares.

Existen cuatro tipos de control de calidad:

- Lingüística

Comprueba la calidad de la traducción, la consistencia del texto y la terminología y su estilo. Incluye la creación y mantenimiento de glosarios y bases de datos terminológicas, creación de guías de estilo y estrategias de edición y corrección de pruebas.

- Técnica

Consiste en la verificación del funcionamiento del sitio, utilizándolo con diferentes sistemas operativos, comprobando cada uno de los archivos, verificación del código, enlaces, etiquetas de HTML, funcionamiento de los formularios y bases de datos, etc.

- Visual

Comprobación del diseño de página y los cuadros de diálogo y textos generados dinámicamente.

- Funcional

Verifica el funcionamiento correcto del sitio localizado en diferentes redes, con distintas bases de datos, navegadores, etc.

Una vez el sitio web cumple todos los requisitos de calidad y funcionamiento, se procederá a su publicación. Una vez publicado, el sitio deberá ser mantenido y actualizado. La solución ideal es contar con un *webmaster* que hable la lengua término para llevar a cabo el mantenimiento diario del sitio. Para llevar a cabo las actualizaciones, habrá que implementar un sistema adaptado al tamaño y tipo de sitio. En la actualidad no existe ningún estándar en el mercado para gestionar el mantenimiento y actualización de sitios localizados y existen miles de herramientas y opciones.

6. Conclusión

Estados Unidos y el Reino Unido ya no son los únicos usuarios de Internet. El aumento de su uso en otros países presenta nuevos retos y requiere nuevas soluciones para ofrecer información en otros idiomas. Este proceso no sólo significa traducción sino que también implica adaptación cultural y la superación de varios problemas técnicos. Se debe adoptar una nueva forma de diseñar y crear sitios web que permita una fácil adaptación a otros idiomas. El proceso de localización de sitios web aún está desarrollándose y nuevas herramientas aparecen a diario en el mercado. En estos momentos cada proyecto de localización es único. Para poder localizar perfectamente un sitio web en varios idiomas diferentes se debe adoptar un enfoque flexible, utilizando diferentes herramientas y métodos basados en un robusto sistema de *workflow*.

Bibliografía

1. DePalma, D.A. "Strategies for Global Sites", Forrester Research Inc, Mayo 1998 y "The eBusiness Report", *Emarketer*, Diciembre 1999.
2. Ebben, S. y Marshall, G. The Localization process: Globalizing your Code and Localizing your Site, <<http://msdn.microsoft.com/workshop/management/intl/locprocess.asp>>. Fecha de la última actualización: 21-04-1999. Fecha de consulta: 6-06-2000.
3. Ebben S. y Marshall G. Designing a Globalized and Localizable Web Site, <<http://msdn.microsoft.com/workshop/management/intl/designloc.asp>>. Fecha de la última actualización: 27-08-1999. Fecha de consulta: 20-06-2000.
4. Esselink, B. (1998), *A Practical Guide to Software Localization*, Amsterdam/Philadelphia: John Benjamin Publishing Company.
5. Esselink, B. (2000), "Localisation Project Models", *Tranfree*, Número 13.
6. ITP Group Technical, Localising Web Sites and Content on the Internet, <<http://www.itp.ie>>. Fecha de la última actualización: 1-08-1999. Fecha de consulta: 21-06-2000.
7. Lingo Systems, (1999) *The Guide to Translation and Localization: Preparing products for foreign markets*. California: IEEE Computer Society.
8. Nielsen, J. (1999). *Designing Web Usability*. Indianapolis: New Riders Publishing. *Computer Dictionary* (1994), Microsoft Press.
9. Roche, M. "Managing Multilingual Web Applications", *16th International Unicode Conference*, Amsterdam, Marzo 2000.
10. Uren, E. et al. (1993). *Software Internationalization and localization: an Introduction*. Nueva York: Van Nostrand Reinhold.