El libro electrónico

¿Qué es el libro electrónico?

El libro electrónico es un formato digital de “libro” que debe seguir unas pautas concretas y especiales de etiquetado para que posteriormente pueda ser utilizado y leído en una serie de dispositivos especialmente diseñados para la lectura de este tipo de formatos.

La principal característica para que un libro electrónico sea tomado como tal es que el texto debe ser libre y fluir, perdiendo todo significado el concepto de “página”.

Esto no quiere decir que por ello tenga que perder su estructura o estilos, que deben ser tan parecidos al original como sea posible, o bien, ser adaptados para su correcta y agradable lectura en formato digital.

El libro electrónico

¿Qué elementos forman un eBook?

Existen una serie de organismos y organizaciones que marcan las pautas de etiquetado que debe seguir un libro electrónico. Conviene ceñirse a éstas normas para obtener un producto estándar.

Estas directrices marcan que un libro electrónico debe tener un OPF (Open Packaging Format) que define cuál es el formato y los elementos que lo forman.

Además debe tener una tabla de navegación o de contenidos donde se describen los contenidos y el índice de éstos (TOC)

Estos dos elementos son de gran importancia porque definen la estructura y el estándar.

El libro electrónico

Un libro electrónico está formado, además de por los elementos antes comentados, por los contenidos. Éstos se guardan en un lenguaje etiquetado estándar que hace posible su lectura correcta por cualquier tipo de dispositivo.

El formato utilizado puede ser XML o XHTML que son variantes mejoradas del HTML original.

Además contiene todas las imágenes que forman parte del libro, incluida la cubierta. Éstas deben ser declaradas y referenciadas en el OPF del libro.

También incluye una hoja de estilos (CSS) donde se definen los estilos que tendrá el libro en cuestión, junto con las fuentes tipográficas que se empleen.

El libro electrónico

Estructura de un libro electrónico

OPF (define la estructura y los elementos que forman el eBook)

TOC (define la tabla de Contenidos y el índice del eBook)

CONTENIDOS (en formato XML o XHTML correctamente etiquetados)

IMÁGENES (incluidos como Elementos sueltos referenciados en el OPF)

CSS (con todos los estilos Utilizados en el eBook)

FUENTES (todas las fuentes tipográficas utilizadas en el eBook)

Formato Compilado (epub)

DISPOSITIVOS DE LECTURA

Formatos y tipos de ebooks

Dispositivos de lectura

Son los dispositivos específicos que van a permitir poder utilizar y leer un libro electrónico. Existen distintos fabricantes y dispositivos y cada día aparecerán nuevos (Sony, Inves, Papire, Kindle, IPhone, iPad, etc.

Software de lectura

Son los programas que nos van a permitir leer los libros electrónicos en dichos dispositivos y marcan la mayoría de las posibilidades y opciones que se pueden hacer (hojas de estilo que se pueden utilizar, funcionalidades, etc). Adobe Digital Edition, Stanza, Mobipocket Reader, etc.

Formatos de libro electrónico

Son los formatos en los que se guarda el libro electrónico. Suelen mantener una estrecha relación con el software de lectura y éste a su vez con el dispositivo. ePUB, mobipocket, PDF reader, etc.

El proceso editorial y el eBook

Problemática actual

Actualmente se pueden establecer dos grandes grupos de libros publicados por una Editorial:

- El fondo de catálogo y sus reediciones
- La novedad

Ambos grupos deben pueden ser considerados de manera distinta pues las posibilidades de trabajo con unos y otros son muy diferentes.

Mientras que el fondo de catálogo sólo suele encontrarse almacenado en formato PDF, la novedad puede tratarse pensando ya que tendrá como destino, al margen de la imprenta, un libro electrónico.

El fondo de Catálogo

El fondo de catálogo

Generalmente se encuentra disponible en formato PDF y es desde este formato desde el que hay que partir para conseguir llegar a una estructura de libro electrónico correctamente formada y etiquetada.

Son distintos los mecanismos y procesos que se pueden seguir para llegar a ese formato de libro electrónico.

La mayor dificultad estriba en poder extraer toda la información desde el PDF puesto que éste es un formato pensado para impresión más que para la reutilización de los contenidos originales.

Existen técnicas y protocolos que permiten etiquetar y reutilizar los PDFs pero por regla general no se utilizan a la hora de generar un PDF cuyo destino sea la imprenta.

Generación de un eBook desde un PDF

Técnicas para convertir un PDF

Para poder extraer la información contenida en un PDF, tanto el texto, como las imágenes como los estilos con los que está diseñado, existen distintos filtros automáticos de exportación, pero la mayoría dan resultados poco satisfactorios, pues al ser un formato pensado para impresión, los datos se encuentran almacenados de forma muy específica y en la exportación el texto suele aparecer desordenado y sin estilos, y las imágenes pierden su ubicación.

Tras la exportación el trabajo de “rehacer” los contenidos y darles formato puede resultar un trabajo muy laborioso que incluya muchos procesos manuales. Prácticamente remaquetar el libro.

La mejor forma de extraer los contenidos es profundizar en el PDF y conseguir extraer toda su estructura completa, incluidos los estilos originales y la ubicación de las imágenes, de una forma automática que ahorre costes y tiempos.

Generación de un eBook desde un PDF

Generación automática de un eBook

1. Se divide el PDF en partes que servirán para determinar la estructura del libro electrónico (que se guardará en el OPF) y el índice (que se guardará en la TOC).

2. Con diferentes procedimientos automáticos programados se extraen los contenidos en formato XML o XHTML. En el proceso hay que tener en cuenta que la separación de palabras con guión por final de línea en el libro electrónico no tienen sentido.

3. Por otro lado se extraen las imágenes u otros elementos que forman parte del libro, junto con su ubicación exacta dentro del libro.

4. Además se extraen los estilos que forman cada parte del libro (indentados, sangrados, párrafos, líneas blancas, etc.) así como los que vienen determinados por los estilos de fuente (itálicas, negritas, versalitas, capitulares, con o sin serif, etc.).

Generación de un eBook desde un PDF

5. Se eliminan elementos no necesarios para el libro electrónico (números de páginas, marcas y cabeceras de impresor, cabeceras de página repetitivas, etc.).

6. Una vez que todos los elementos se han extraído, éstos quedan almacenados en un repositorio perfectamente etiquetado que nos permitirá generar un libro electrónico en cualquier formato.

7. Ese repositorio perfectamente etiquetado deber ser el objetivo a perseguir porque nos va a permitir generar cualquier formato de libro electrónico en un futuro (epub, mobipocket, PDF reader, etc.) o bien poder reutilizar esos contenidos con otros fines (Internet, dispositivos móviles, TV a la carta, etc.) de una manera muy sencilla y sin tener que realizar otros procesos que resulten muy costosos.

Bastará con seleccionar los contenidos que queremos utilizar en cada caso y definir una hoja de estilos apropiada.

Generación de un eBook desde un PDF

8. En función del formato de libro electrónico que se vaya a utilizar, y el dispositivo al que vaya a ir destinado, todos esos contenidos almacenados en formato XML o HTML, más la estructura de elementos (OPF), índices (TOC), imágenes, estilos y fuentes deben ser compilados en un formato u otro.

9. Actualmente el formato más estándar es el ePUB, puesto que puede ser utilizado por la gran mayoría de dispositivos más populares (Inves, Sony, Papire, con Adobe Ditital Edition, iPhone –mediante Stanza–). El dispositivo de Amazon, Kindle, utiliza actualmente mobipocket (al margen de otros formatos propietarios). En general todos pueden leer PDFs, pero el PDF no es un formato adecuado porque mantiene la estructura de página y dificulta la lectura.

10. Una vez que tenemos los contenidos perfectamente estructurados y etiquetados el paso a cualquiera de estos formatos es prácticamente automático.

Generación de un eBook desde un PDF

Para la conversión de libros de fondo catálogo en formato PDF se puede establecer la siguiente clasificación, en función de la complejidad de sus elementos (tablas, notas, imágenes, fuentes, etc.) y de los estilos presentes en el libro:

1. Libros Sencillos
2. Libros Normales
3. Libros Difíciles
4. Libros Especiales

De manera estimada se puede decir que de manera genérica el catálogo de una editorial puede seguir el siguiente reparto de porcentajes por tipo de libro:

- Un 45% de libros sencillos
- Un 40% de libros normales
- Un 10% de libros difíciles
- Un 5% de libros especiales

Como es lógico, los costes de conversión varían en función de la dificultad del libro.

Generación de un eBook desde un PDF

Para llevar a cabo la conversión hay que realizar las siguientes tareas:

1. Sincronización de lotes / libros
2. Troceo de PDFs en capítulos
3. Generación de estructura e índices
4. Recepción de metadatos (XML Onix)
5. Conversión
6. Aplicación de estilos.
7. Validación técnica.
8. Revisión editorial
9. Validación en dispositivos de lectura.
10. Sincronización final de los productos obtenidos (epub, recursos, XML ONIX)

Generación de eBOOKS - Novedad

La Novedad

A la hora de generar libros electrónicos partiendo de libros que todavía no han sido creados, pudiendo participar activamente en el proceso editorial en su creación, las posibilidades son mucho mayores.

Uno de las principales problemas a la hora de conseguir un eBook partiendo de un PDF surge de conseguir mantener los estilos, fuentes y colocación de los textos del original.

Este problema aparece porque los libros originalmente no se crearon pensando en un reaprovechamiento de los activos editoriales en formato digital, sino para su uso exclusivo en impresión.

Modificar mínimamente alguna de estas técnicas permite que los tiempos y costes de conversión se reduzcan notablemente.

Generación de eBOOKS - Novedad

Un correcto etiquetado de los contenidos, respetar los estilos a lo largo de todo el libro, marcar cualquier elemento no ordinario (como las tablas, los pies de figura, las notas, etc.) y ceñirse a unas hojas de estilo correctas reducen enormemente los tiempos y los costes de conversión.

Siguiendo una serie de pautas y prácticas específicas para el libro digital, a la vez que se maqueta el libro para su producción por impresión, podemos conseguir generar ambos productos simultáneamente con un incremento de costes muy inferior al que tendríamos en caso de no seguirlas.

Además es muy recomendable pensar en el soporte al que va destinado el libro electrónico, porque no todos los libros son válidos para dicho soporte. También se debe replantear que ciertos estilos y estructuras deben modificarse cuando el libro va a ser presentado en formato digital.

Consideraciones generales

Independientemente de si el libro electrónico se va a generar a partir de un libro almacenado en PDF o a partir de una novedad, en la que se puedan especificar las características necesarias antes comentadas para su mejor conversión, es necesario tener en cuenta una serie de consideraciones generales aplicables en conjunto.

1. Utilización de fuentes incrustadas.
2. Utilización de hojas de estilo específicas para el libro electrónico.
3. Definición de protocolos en la conversión.
4. Posibles defectos o malas prácticas en la generación original de los PDF

Consideraciones generales

Utilización de fuentes incrustadas

Es un importante debate planteado actualmente. Los libros electrónicos ganan en riqueza al incrustar fuentes y hoy en día es posible hacerlo en la mayoría de dispositivos.

Además se consiguen eliminar muchos problemas en la conversión puesto que muchos de los estilos vienen determinados por las propias fuentes, estilos que al prescindir de ellos pueden perderse obligando a realizar más procesos para convertir un libro, subiendo de esta manera tiempos y costes.

Los principales fabricantes como Sony, o proveedores como Adobe, recomiendan el uso de fuentes.

La parte negativa, lo que puede establecer el debate, son los derechos de distribución, aunque actualmente existen ya fórmulas comerciales que contemplan estos derechos.

Las fuentes siempre deben ser Open Type para garantizar una correcta distribución en cualquier tipo de dispositivo.

Consideraciones generales

Utilización de hojas de estilo

Los libros electrónico, de forma similar a como lo hacen las páginas web, definen la forma que tiene de mostrarse mediante unas hojas de estilo que marcan la presentación.

Existen dos opciones a la hora de desarrollar estas hojas de estilo. La primera es intentando emular al libro original lo máximo posible. De esta forma cada libro tendrá una hoja de estilo propia generada a partir del libro original.

La segunda, y más recomendable, es generar una hoja estilo (o varias si fuera necesario, por sello, por colección), y que los libros electrónicos se adapten a dichas hojas de estilo.

Trabajar de esta manera permite que libro electrónico tenga una serie de definiciones de estilo específicas para su lectura en pantalla que hagan que dicha experiencia sea mucho más agradable.

La definición de hojas de estilo es un punto fundamental a la hora de afrontar los procesos de conversión y deben estar establecidas antes de comenzar los mismos.

Consideraciones generales

Definición de protocolos

A la hora de iniciar los procesos de conversión es necesario establecer una serie de protocolos entre la editorial y el conversor que definan cómo se van a generar los libros electrónicos.

Una buena definición ahorrara muchos problemas posteriores y significará un ahorro de costes y tiempos.

Establecer cómo estará definida la TOC, que elementos se incluirán en el índice o no. Definir la plantilla de página de créditos, página corporativa si la hubiera, que elementos existentes en el libro en papel aparecen en el libro electrónico y cuales no. Definir como se trabajará con las notas de pie de página, del editor o del traductor, si en fichero aparte, con hiperenlaces o integradas en el texto, etc.

Todas estas definiciones, si se establecen de manera correcta a priori, evitarán tener que generar los libros varias veces, evitando costes y rebajando los tiempos de conversión.

Consideraciones generales

Malas prácticas en el original en PDF

A la hora de maquetar los libros destinados a papel, en el proceso de autoedición, hay muchas prácticas que se siguen habitualmente y que no tienen ninguna consecuencia en la impresión, pero a la hora de generar los libros electrónicos si pueden generar serios problemas.

La utilización de falsas itálicas (inclinando el texto en lugar de utilizando una fuente itálica). Abusar del Kerning o el Tracking para cerrar páginas (puede juntar palabras en el PDF aunque visualmente estén en el límite). La utilización de falsas versalitas (en lugar de utilizando el estilo “smallcaps” o “versalita”, reduciendo el tamaño de la fuente). Incrustar subconjuntos de fuentes en lugar de conjuntos, (puede ocasionar que determinados caracteres se pierdan), etc.

Todas estas técnicas pueden generar problemas a la hora de extraer la información del PDF, teniendo como resultado que los originales en PDF que guardamos no sean correctos para su posterior utilización.

Conclusiones finales

Definir los protocolos y pautas correctas, tanto desde el punto de vista editorial, como en los procesos de conversión, y que éstos estén perfectamente definidos y acordados entre el proveedor y el editor es fundamental para una correcta conversión.

Seguir una serie de pautas y técnicas correctas a la hora de generar los libros, ya sean para formato en papel o para formato electrónico, significará un ahorro importante de costes y tiempos en los procesos de conversión.

Definir o redefinir las políticas editoriales a la hora de afrontar la conversión de libros a formato electrónico, en aquellos aspectos que influyan a la hora de disponer de libros en este nuevo medio es fundamental.

Los formatos y los procesos cambian y tenemos que adaptarnos a estas nuevas necesidades, eliminando o modificando pautas y comportamientos que dificulten esta adaptación al nuevo medio.

Newcomlab

Gestión y desarrollo de negocios en Internet

www.newcomlab.com

CONVERSIÓN DE LIBROS ELECTRÓNICOS

Raúl Pereda Ruíz
rpereda@newcomlab.com
+34 91 728 06 26 (Ext. 18)
687 768 477