

Trabajando con Docbook

Por Ismael Olea <ismael@olea.org>

Una conferencia CACLE



Objetivos

- Conocer la DTD Docbook para crear documentación informática.
- Conocer las herramientas, el entorno de trabajo de docbook y los problemas más típicos.

Los documentos tienen estructura

Se presupone conocer los principios de la estructuración de documentos.

Presentación de Docbook

- DTD técnica de informática
- libros, artículos, páginas de manual y colecciones de los mismos
- diseñado por la industria
- actualizado
- SGML/XML
- DSSSL/XSL
- estandarizado en todos los proyectos de soft libre importantes
- incluido en todas las distribuciones modernas

Recursos de trabajo mínimos

- Manual de referencia: Docbook: The Definitive Guide
- jade/openjade
- juego de dtDs
- juegos de caracteres ISO
- docbook utils
- jadetex + TeX («opcional» para imprimir)

Paquetes para RH 7.*

Básicos:

- docbook-dtds
- docbook-style-dsssl
- docbook-utils
- docbook-utils-pdf
- jadetex
- tetex-*
- openjade
- sgml-common
- passivetex
- docbook-style-xsl
- xmlto

Opcional:

- psgml

Paquetes para Debian Woody/LineX

Método básico:

- `apt-get install docbook-utils`
- <http://tldp.org/authors/tools/ldp.dsl>

Configuraciones: catálogos

DOCTYPE

- tipo de documento
- PUBLIC|PRIVATE
- Id de la versión de la DTD

Los catálogos son los directorios de recursos SGML/XML para su localización automática por parte del sistema. Deben funcionar de forma transparente para el usuario. En proceso de estandarización.

- /etc/sgml
- /etc/xml

Configuraciones relacionadas con TeX

Para obtener los mejores resultados de impresión posibles, antes es preciso configurar o comprobar algunos puntos.

Se usa TeX para la obtención de formatos de impresión PDF/PS. Jadedtex/pdfjadedtex es un intérprete TeX de macros para jade/openjade.

- Patrones de guionado de TeX
 1. # texconfig
 2. ->HYPHEN
 3. ->pdflatex
 4. descomentar «%spanish sphyph.tex» (o los idiomas que tercién)
 5. salir de texconfig
- Jadedtex
 - Configuración de los buffers de TeX
 - Configuración del idioma español: jadedtex.dtx
 - `\RequirePackage[german,frenchb,english]{babel}[1997/01/23]`

Estudio de un documento de ejemplo

A continuación examinaremos un ejemplo completo y funcional explicando cada parte paso a paso.

`apannao.xml`

Usando el manual de referencia

Presentación de The Definitive Guide.

- estructura del libro:
 - introducción
 - referencia
 - marcas
 - entidades parámetro
 - entidades de juegos de caracteres
 - apéndices
- Explicación rápida de los juegos de caracteres (por qué hay que tenerlos en cuenta).
- Explicación de las secciones de una página de referencia de un elemento.
- Explicación de metodología propuesta de uso.

Trabajando: definición de entidades y descomposición en varios ficheros fuente

Las entidades tienen varios usos diferentes. Los más corrientes son:

Codificación de caracteres

 sirven para codificar, en ascii, caracteres nacionales o de difícil representación.

Entidades generales internas

 sirven para crear acrónimos o macros.

Entidades generales externas

 se usan para partir el código fuente en varios ficheros.

Encontrarán información más detallada en *The Definitive Guide*.

Trabajando: equivalencias prácticas entre Docbook SGML y Docbook XML

- Con alguna rara excepción ambas DTD son iguales.
- Como XML es compatible con SGML es posible usar el ciclo de producción SGML sin inconveniente (la vía DSSSL).

Trabajando: diferencias prácticas entre Docbook SGML y Docbook XML

- El sufijo :-)
- El código XML permitiría usar las herramientas XSL (no estudiadas en este curso).
- Con XML los elementos (las marcas) siempre estarán en minúsculas.
- Con XML siempre se escribirá el contenido de los atributos entre comillas.
- Con XML las marcas que no necesitan ser cerradas se escriben <marca/>.
- Con XML la primera línea siempre será una declaración XML explicando la versión de XML y el juego de caracteres.

Trabajando: usando un corrector ortográfico

```
aspell --lang=español -c -H fichero.xml
```

Trabajando: generando formatos de reproducción/impresión

Obtener PDF:

```
docbook2pdf apannao.xml
```

Obtener un documento HTML descompuesto en ficheros:

```
docbook2html -u apannao.xml
```

Obtener HTML:

```
docbook2html apannao.xml
```

Otros filtros: `docbook2dvi` `docbook2man` `docbook2ps` `docbook2tex` `docbook2txt`
`docbook2rtf` `docbook2texi`.

Trabajando: personalización de DTD

SGML y XML están concebidos para poder hacer modificaciones o extensiones a las DTD.

En la práctica esta necesidad es muy rara. Además, al modificar la DTD perdemos facilidades de portabilidad entre sistemas y exige la ampliación o modificación de las hojas de estilo.

Por todo ello no se considerará en este curso.

Trabajando: personalización de hojas de estilo DSSSL

Es posible personalizar el comportamiento de las hojas de estilo sin modificarlas usando un «driver» DSSSL.

Este «driver» sirve para *matizar* a las hojas actuales y deben incluir una referencia a las mismas para que el procesador DSSSL las localice.

El driver se invocaría, por ejemplo:

```
docbook2pdf -d ldp.dsl#print fichero.xml
```

Trabajando: personalización de hojas de estilo XSL

Dado que el ciclo de producción XSL no está tan maduro como el DSSSL no consideraremos en este curso la personalización de sus hojas de estilo.

Trabajando: personalización de hojas de estilo CSS

Trabajando: uso de otras herramientas

Es posible usar herramientas que aumentan la productividad. Por ejemplo:

CVS

Control de versiones y trabajo en grupo.

Make

Control del código fuente.

Trabajando: experiencia: DocBook vs OpenOffice

Un ejemplo real de uso en grupo partiendo de cero, contada por J.Manrique Lopez.

- El director del grupo, partiendo de un desconocimiento técnico total, decide usar DocBook para el documento del anteproyecto.
- La coordinación del grupo, 13 personas, fue muy complicada, (mucha gente y muy acostumbrados al formato DOC).
- La única queja en la entrega, PDF, fue sobre el tamaño de la letra, el desconocimiento impidió crear un driver .dsl a tiempo.
- Para la segunda parte, a pesar de tener el driver .dsl y por problemas ajenos al desarrollo del proyecto, se debió cambiar a otra opción y el director del grupo optó por OpenOffice.
- Hubo quejas de todo tipo:
 - Falta de enumeración en los títulos.
 - No hay manera de dejar 2 tablas iguales: cada miembro del equipo las hizo como le vino en gana [...] y eso que se había acordado que se hiciesen con

hoja de cálculo.

Los defectos fueron atribuidos a haber utilizado Software Libre, como si con Word no hubiera podido hacer lo mismo (de hecho Word necesita Acrobat para exportar PDF).

Lo más doloroso fue oír a un profesor de universidad comentar: *Parafraseando al Ministro de Tecnología sobre estos temas (SL), los experimentos en casa y con gaseosa.*

Y matizó: *Con lo bien que os había quedado el Anteproyecto.*

Complicaciones más comunes: errata Debian/Linex

Una errata en las hojas de estilo de la actual versión de Linex impide que algunas imágenes se incluyan en los ficheros pdf de salida. Esto puede resolverse fácilmente usando un «driver», como por ejemplo: <http://tldp.org/authors/tools/ldp.dsl>. El modo de uso sería:

```
docbook2pdf -d ldp.dsl#print fichero.xml
```

Complicaciones más comunes: juegos de caracteres

Hasta ahora el juego de caracteres predeterminado en Linux para el español ha sido Latin1 (iso-8859-1). La nueva versión adaptada al Euro es Latin9 (iso-8859-15). Ambos juegos son muy parecidos. Como el sistema funciona siempre funciona con uno y otro no es fácil que se puedan mezclar.

Sin embargo el juego predeterminado en la tecnología XML es UTF-8 y es probable que nos encontremos muchos ficheros así codificados. A la hora de modificar un fichero XML hay que tener presente que la codificación del fichero y la que usa la aplicación coincidan. Si no es así se pueden hacer transformaciones con programas como el analizador xmllint o **recode**, un poderoso conversor de codificaciones.

Complicaciones más comunes: usando ilustraciones en PS/PDF

Como hemos visto TeX es el motor de composición para los formatos de reproducción «impresa».

Hay dos motores TeX, el clásico que crea ficheros PostScript (y el intermedio DVI) y PDFTeX que crea PDF.

Curiosamente ambos motores no manejan los mismos formatos gráficos para ilustraciones. El lío aumenta al tener en cuenta la salida vía web.

Hay varias formas de sortear este problema pero la más sencilla es optar siempre por la salida PDF y HTML y utilizar los formatos gráficos PNG y JPG. Y no especificar nunca el atributo **format**.

Complicaciones más comunes: usando ilustraciones en HTML

Ilustraciones

Algunas versiones de docbook-utils varían de comportamiento al crear documentos HTML descompuestos en varios ficheros. Así en unos casos todos los ficheros html se crearán en el directorio actual y en otros lo hará en un subdirectorio. En este segundo caso será necesario copiar las imágenes a una ruta relativa igual a la que se introdujo en el código fuente para su correcta visualización en el navegador.

Adornos de estilo

Según la configuración de las hojas de estilo éstas podrán usar adornos con forma de fichero gráfico. Si usa esta opción deberá copiar a mano estas imágenes (en mi caso en `/usr/share/sgml/docbook/dsssl-stylesheets-1.74b/images/`) al subdirectorio adecuado.

Editores estructurados

El código Docbook es bastante prolijo. Puede que prefiera usar un editor estructurado más avanzado y cómodo. Algunas alternativas libres o gratis:

vim

dispone de un modo de coloreado para XML, pero no asiste a la escritura.

jedit

Es un editor java, portable, con modos de edición de XML, XSLT, etc.

emacs+psgml

Es un muy poderoso editor estructurado. Una versión más acabada de psgml es psgmlx.

xxe

Es un editor estructurado en java.

Referencias

Docbook: The Definitive Guide

<http://www.oasis-open.org/docbook/documentation/reference/>

Referencia principal

<http://www.docbook.org>

Docbook Open Repository Project:

<http://docbook.sourceforge.net>

Otro relacionado de Norman Walsh

<http://www.nwalsh.com>

Tutorial de Docbook (usado en este tutorial):

<http://olea.org/conferencias/doc-conf-tutorial-docbook/slide001.html>

Otra presentación sobre Docbook

<http://es.tldp.org/Presentaciones/0000otras/conf-jdavila1/html/index.html>

Tutorial en español

<http://es.tldp.org/Tutoriales/DOCBOOK/doctut/single-html/dbktut.html>

Docbook Howto (inglés):

<http://www.ibiblio.org/godoy/sgml/docbook/howto/index.html>

Foros de ayuda

Listas oficiales de consulta sobre Docbook y aplicaciones Docbook (inglés)

<http://www.oasis-open.org/docbook/maillinglist/index.shtml>

Lista de ayuda en español

<http://listas.hispalinux.es/mailman/listinfo/docbook-ayuda/>

Preguntas

Atrévase a preguntar todo lo que siempre quiso saber sobre Docbook

Despedida y cierre

¡Esto es to... esto es to... esto es todo amigos!