Bibliotecas electrónicas: tecnología de documentación

Ismael Olea
<mailto:olea@hispalinux.es> 1999-09-18

1 Introducción: del LDP a la biblioteca inteligente

A lo largo de este tutorial conoceremos algunas de las tecnologías necesarias para la puesta en marcha de proyectos documentales. Estas ideas han sido observadas desde la perspectiva de la experiencia a lo largo de mi contacto con los proyectos de documentación LDP, LuCAS y LDP-es en general.

Lo que se ofrece es una idea de cómo debe ser la próxima generación de proyectos de documentación GNU/Linux en concreto y de bibliotecas electrónicas en general.

Un proyecto de documentación técnica-científica informatizada necesita de las siguientes tecnologías:

1. documentos estructurados y asistentes a la creación;
2. libro de estilo;
3. sistema de control de responsabilidad distribuida;
4. sistema automático de publicación.

1.1 Documentos estructurados y asistentes a la creación

El código fuente debe estar basado en un sistema abierto de documentos estructurados basado en las tecnologías SGML/XML.

Ventajas (expuestas también en la parte dedicada a SGML/XML):

- Marcación descriptiva versus marcación procedimental.
- El autor se centra en la labor de escritura y creación, sin distraerse con asuntos que no son relevantes en el proceso de autoría.
- El documento es independiente de la plataforma, facilitando su manejo, almacenado, consulta y proceso.
- Sencillez de edición y fácil manejo con cualquier clase de herramientas de edición de textos.
- Apoyado por un adecuado sistema de representación mejora la calidad final de los documentos y unifica el uso de las convenciones de estilo de cada idioma y área de conocimiento.
- Gracias al riguroso diseño también se facilita la automatización del proceso y el intercambio de documentos por medios informáticos.

En cuanto a las herramientas que asistan a la creación:

- Biblioteca de DTD, definiciones de tipos de documentos.
- Asistentes para la creación de DTD.

- Hojas de estilo para la transformación a otros DTD y a formatos de visualización.
- Editores estructurados.
- Editores de texto.
- Tesauros.
- Corrector ortográfico.
- Asistente para la creación de meta-datos.
- Control de versiones.
- Tecnologías y herramientas de representación electrónicas y en papel.

1.2 Libro de estilo

El libro de estilo es una herramienta absolutamente indispensable para todos aquellos que sin ser profesionales de la documentación necesitan crear textos destinados al público. Así pues el libro de estilo es la guía y oráculo que debe ayudar a resolver todas las dudas e inquietudes del autor.

Así debería incluir al menos las siguientes secciones:

Convenios de escritura
: donde se normalice el uso del estilo literario dentro del documento o de la colección de los mismos.

Vocabulario
: donde se fije una terminología uniforme a fin de no despistar al lector.

Glosario de traducción
: algo que parece imprescindible en el caso de la documentación informática en el que tanta importancia tiene el vocabulario de la jerga en inglés. Es obvio que debe estar perfectamente sincronizado con el vocabulario o que ambos formen una unidad.

Convenios tipográficos
: que reflejen la norma y convenios del idioma usado e instruyan su aplicación en la tecnología de documentos elegida.

Recursos de ayuda al autor/traductor
: bibliografía electrónica e impresa de interés.

1.3 Sistema de control de responsabilidad distribuida

Las labores que debe favorecer este sistema son:

- Control de versiones.
- Creación cooperativa.
- Actualización de los documentos.
- Control de calidad.

Con «responsabilidad distribuida» quiero referirme al modo en el que se desarrollan los proyectos tipo *bazar*, en los cuales la responsabilidad de las tareas las adquieren voluntariamente los colaboradores en función de sus posibiliades y deseos; paralelamente el proyecto es dirigido/supervisado por un equipo más o menos reducido de líderes. Así es importante que las herramientas agilicen las tareas de coordinación al máximo liberando a los líderes del tiempo dedicado a las labores rutinarias.

1.4 Sistema automático de publicación

Actualmente estoy trabajando en la especificación de los requerimientos a exigir a esta clase de tecnologías. Es lo que he llamado la «arquitectura **Donantonio**».

Los planteamientos básicos de **Donantonio** son:

- Basado en soft libre, preferiblemente GPL o compatible.
- Especificaciones abiertas y perfectamente documentadas.
- Basado en verdaderos estándares abiertos como RFC, W3C e ISO.
- Sencillez, modularidad y flexibilidad.

Los objetivos del diseño son:

- publicación automática;
- publicación distribuida a lo largo de la Red;
- responsabilidad distribuida de la publicación;
- organización de vistas —de usuario y de sistema— automáticas y programables;
- facilidades de replicación;
- facilidades de actualización automática;
- relaciones flexibles entre agentes **Donantonio**.

Elementos necesarios (entre otros):

- declaración de metadatos conforme a un esquema normalizado (descripción);
- el fichero como unidad atómica de transferencia de datos;
- protocolos de transferencia de ficheros internet;
- mecanismo de consulta sencillo y flexible (por ahora, SQL).

2 Contenidos de este tutorial

A lo largo de este tutorial se mostrarán técnicas que dan solución a parte de las necesidades del primer punto y al cuarto.

2.1 Documentos estructurados.

Una introducción a las tecnologías de documentos estructurados está disponible en la conferencia «SGML para autores» que puede encontrarse en <http://lucas.hispalinux.es/Presentaciones/campus-party98/conf-olea-cp98.tgz>.

2.2 Sistema automático de publicación.

En la actualidad estoy diseñando el borrador de una arquitectura para el diseño y operación de esta clase de sistemas. Además participo en el diseño de una aplicación bibliográfica que asista a la publicación de recursos en **Metalab** con la vista puesta en un diseño basado en **Donantonio**.

También tengo previsto que este esfuerzo, junto a otros que al menos partirán inicialmente de la comunidad hispana, forme parte del próximo proyecto LuCAS I+D.

Por lo pronto ya están en desarrollo los siguientes trabajos:

LDP-core,

esquema bibliográfico para clasificar los documentos del LDP. <TACATACA>.

ldpCoreMetaML.dtd,

LDP-core expresado como DTD XML. <http://casbah.org/cgi-bin/cvsweb.cgi/XML/DTDs/ldpCoreMetaML.dtd>

Donantonio,

una arquitectura para publicación distribuida automática.

3 Glosario

DC

<ftp://sunsite.rediris.es/docs/rfc/24xx/2413> <http://purl.org/DC/>

DocBook

<http://www.docbook.org>

DTD

TACATACA

GNU

<http://www.gnu.org> <http://www.fsf.org>

Glosario de LuCAS

<http://LuCAS.HispaLinux.ES/Otros/glosario/Glosario/glosario.html>

GPL

TACATACA

Insflug

TACATACA <http://www.insflug.org/>

ISO

International Standars Organization

LDP

<http://www.linuxdoc.org>

LDP-es

<http://LuCAS.HispaLinux.ES>

LuCAS

<http://LuCAS.HispaLinux.ES>

Metalab

<http://metalab.unc.edu>

Ofelia

<http://www.insflug.org/ldp/>

RDF

<http://www.w3.org/Metadata/RDF>

RFC

<ftp://sunsite.rediris.es/docs/rfc/>

SQL

TACATACA

URL

Uniform Resource Locators <ftp://sunsite.rediris.es/docs/rfc/17xx/1738>

SGML

TACATACA

XML

<http://www.w3.org/XML/>

W3C

<http://w3.org/>

4 Bibliografía

Además podrá encontrar material actualizado de este tutorial y otros relacionados en

- <http://www.hispalinux.es/~olea>
- <http://lucas.hispalinux.es/htmls/presentaciones.html>