

40

GESTIÓN SEMÁNTICA DE CONTENIDOS EN “EL PORTAL DEL CIUDADANO”

Juan A. Prieto
Gerente área de Tecnología
XimetriX network thoughts, SL

1. EL PORTAL DEL CIUDADANO

El Ministerio de Administraciones Públicas gestiona y publica El Portal del Ciudadano, www.administracion.es, un punto virtual de encuentro entre el ciudadano y las Administraciones Públicas formado por más de 20.000 documentos y un millón de enlaces. En el Portal del Ciudadano tiene cabida información de interés público generada por las diferentes instituciones públicas existentes a nivel nacional. Así, el rango temático es muy amplio, siendo posible encontrar desde una guía para renovar el carné de identidad hasta complejos formularios para la búsqueda de leyes o información actualizada diariamente sobre becas, licitaciones, ofertas de empleo público, etc.

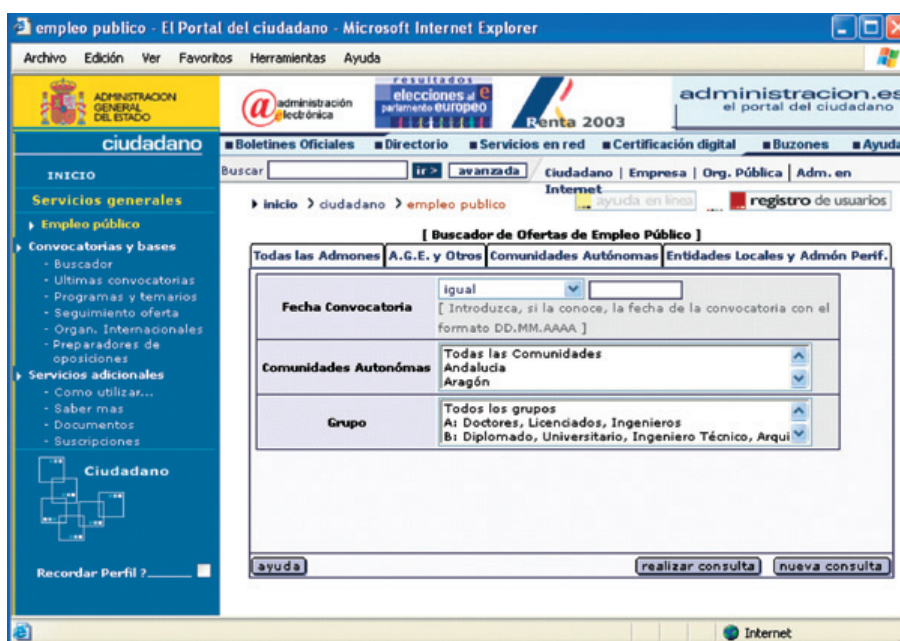


Figura 1: El Portal del Ciudadano

En la práctica, este compromiso se traduce en la necesidad de gestionar contenidos y servicios provenientes de múltiples fuentes, y por tanto heterogéneas, de información para su publicación en diversos idiomas –catalán, eusquera, gallego, valenciano y castellano para el portal en sí; inglés, alemán y francés para la sección internacional del portal–, sobre un rango creciente de canales de visualización y de formatos electrónicos, y verificando las normas de accesibilidad WAI (Web Accessibility Initiative) del consorcio Web (W3C).

El repositorio de información que constituye el Portal del Ciudadano es muy variado, formándolo tanto elementos individuales con una alta rotación y variación en el tiempo (una beca, una licitación, una oferta de empleo público, ...), como documentos ya finales y relativamente estables (una ley, la constitución, etc.).

Otro aspecto importante en relación con la gran variedad temática del Portal es la estructura de la información, de tipo no lineal, en base a relaciones multi-dimensionales. Esta estructura debe ser reflejada en la navegación, garantizando múltiples puntos de acceso a cada elemento de

información y permitiendo la personalización del portal para distintos contextos, perfiles y métodos de exploración de la información publicada. La Figura siguiente ilustra este aspecto:

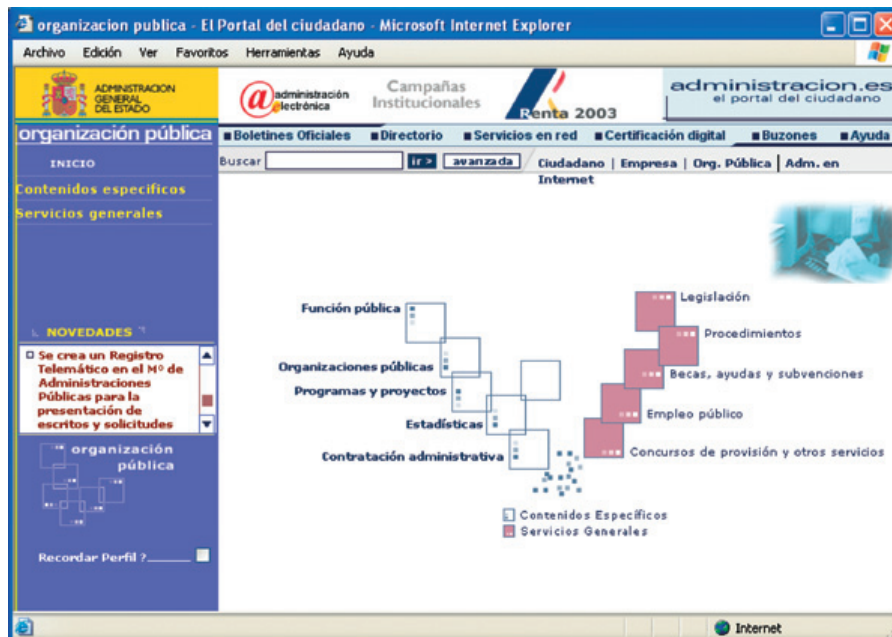


Figura 2: El Portal del Ciudadano, estructuras de navegación

2. GESTIÓN SEMÁNTICA DE CONTENIDOS Y SERVICIOS

El paradigma de la Web Semántica asocia, en un dominio concreto de conocimiento, un significado unívoco a los elementos de información que forman los contenidos [Bern01]. Su virtud principal estriba en el uso de ese significado por los sistemas de información, lo que permite aumentar el grado de interoperabilidad entre aplicaciones, facilitando por ejemplo la agregación y sindicación de información mediante la localización automática de las fuentes y de los destinos de la información.

XimetriX network thoughts (www.ximetrix.com) ha desarrollado ximDEX, un entorno para la adquisición, gestión, transformación y publicación de contenidos, en formatos electrónicos, que asocia un significado bien definido a los elementos de información que forman los contenidos [Pri04].

El Gestor de Contenidos ximDEX¹ permite a los usuarios crear libremente los formatos que estructuran cada tipo de documento (noticia, documento genérico, ley, beca, ...), vinculando un significado preciso (título de noticia, entrada, fecha de inicio, ...) a cada elemento de información individual que forma el documento o servicio.

La estructuración semántica de los contenidos proporcionada por ximDEX constituye un punto clave en el proceso de abstracción y simplificación de los elementos de información y de las relaciones existentes entre ellos, lo que: i) permite una plena separación entre los contenidos y sus capas de presentación [Zapt03, Athe02], ii) facilita la automatización de las tareas de ges-

ción de los contenidos, como, por ejemplo, el cumplimiento de normas de accesibilidad, y iii) permite acercar al personal no técnico las tareas que habitualmente son realizadas por perfiles especializados en labores de programación de sistemas.

El proceso de abstracción en la representación de los elementos que forman los contenidos y servicios que impulsa ximDEX fomenta una plena separación respecto a la presentación, respecto al modelo de datos subyacente y respecto a la lógica de negocio. Esta característica de ximDEX permite simplificar y automatizar todo el proceso de creación y mantenimiento de los contenidos y servicios y, lo que es aún más significativo, reduce la dependencia de los mismos respecto a la tecnología de explotación utilizada (java, php, .net).

De este modo, la creación de un nuevo canal de información (PDA, móvil, etc.), de un nuevo producto para la explotación dinámica de información almacenada, el cumplimiento de normas específicas de publicación para accesibilidad, la agregación de información desde fuentes remotas o la creación de una versión de un portal dirigida a un perfil específico (estudiante, economista, etc.) son tareas puestas al alcance del personal que redacta y organiza los contenidos para su realización rápida, sencilla y libre de errores.

3. GESTIÓN DE CONTENIDOS Y SERVICIOS CON XIMDEX

El entorno ximDEX es un sistema automático de adquisición, transformación y publicación de contenidos y servicios en formatos electrónicos. Los elementos que forman los contenidos y servicios son creados y editados mediante un entorno gráfico tipo wysiwyg, encargándose el entorno de asociar un significado y estructura bien determinados en base a formatos XML definidos con total libertad por los gestores de la información.

El entorno ximDEX permite la automatización de la publicación de contenidos y servicios en los formatos típicos de Internet, ampliando el concepto tradicional de contenidos para englobar además a los componentes encargados de la explotación dinámica. Así, el entorno facilita la creación y gestión de todos los contenidos y servicios, tanto estáticos como dinámicos, transformándolos a los formatos finales a publicar sobre los servidores de contenidos y aplicaciones necesarios.

Consecuencia directa de las capacidades de abstracción que potencia ximDEX, el gestor permite reducir la dependencia tecnológica de los contenidos respecto a las diversas tecnologías de explotación (j2ee, php, .net, ...) y servidores de aplicación existentes. Así, se eliminan las imposiciones tecnológicas en la explotación de los contenidos y servicios por parte del gestor, aumentando en paralelo las capacidades de reciclaje de lógica de negocio ya existente.

De hecho, esta filosofía permite al entorno publicar un mismo servicio de forma simultánea en servidores de aplicación de naturaleza diferente, lo que: i) facilita la migración entre tecnologías, ii) potencia la interoperabilidad entre componentes, y iii) permite aumentar la escalabilidad de los contenidos y servicios.

Desde el punto de vista de desarrollo de los contenidos y servicios, la potenciación de la abstracción respecto a la tecnología final permite una reutilización real de los elementos que los forman, lo que simplifica y facilita su gestión y posibilita el desarrollo de productos dinámicos para la explotación de la información, por parte del personal que habitualmente crea los contenidos en el sistema.

En referencia al tratamiento de idiomas, el Gestor de Contenidos ximDEX posibilita que cada documento tenga un número de idiomas asociado distinto, enlazar una versión idiomática de un documento con otro documento, gestionar dos documentos enlazados de forma concurrente (en fechas de publicación y flujo de trabajo), separar la gestión idiomática de la parte estructural de los documentos (menús, logotipos, apariencia) del contenido propiamente dicho, entre otras funcionalidades.

3.1. Documentos estructurados y anotaciones contextuales

El entorno ximDEX permite de forma sencilla e intuitiva la edición e incorporación de contenidos, garantizándose simultáneamente que los elementos de información se representen de forma estructurada en los formatos XML asociados a cada tipo de documento, formatos libremente definidos por los usuarios con los privilegios adecuados en el sistema.

En la línea de simplificación requerida para el acercamiento del entorno a personal no especializado, el entorno ximDEX proporciona un módulo de importación desde herramientas ofimáticas comunes (MS office, por ejemplo) y un editor wysiwyg para XML, lo que permite la edición de los documentos estructurados sin conocer las etiquetas o propiedades que lo componen, visualizándose simultáneamente el estilo final del documento.

La Figura siguiente muestra el entorno ximDEX durante la edición de un documento XML:

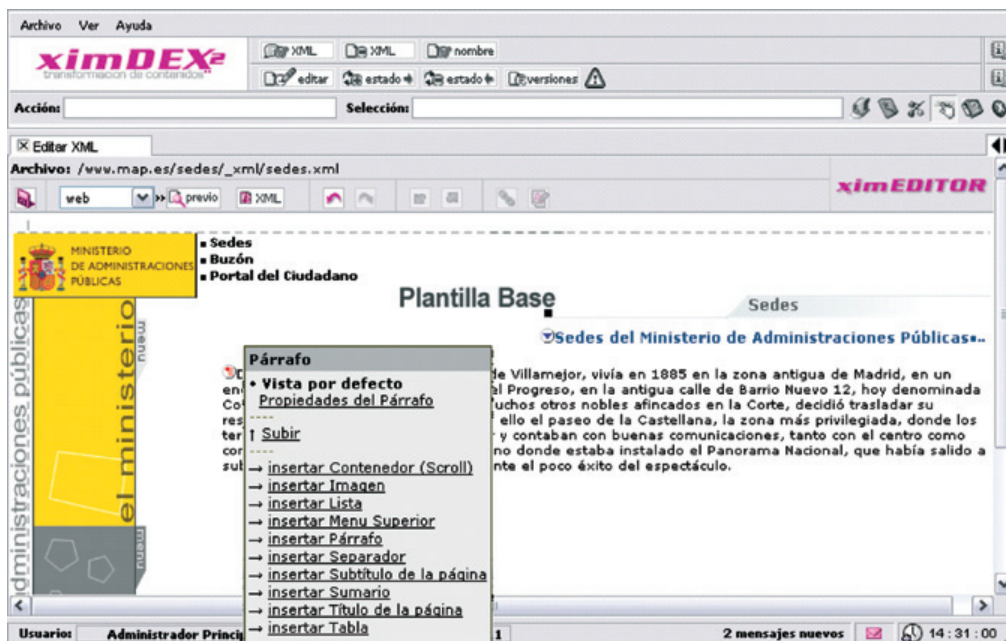


Figura 3: Edición wysiwyg de un documento XML en ximDEX.

Desde el mismo editor se accede también a las propiedades que constituye cada elemento estructurado, proceso que se ilustra en la figura siguiente:

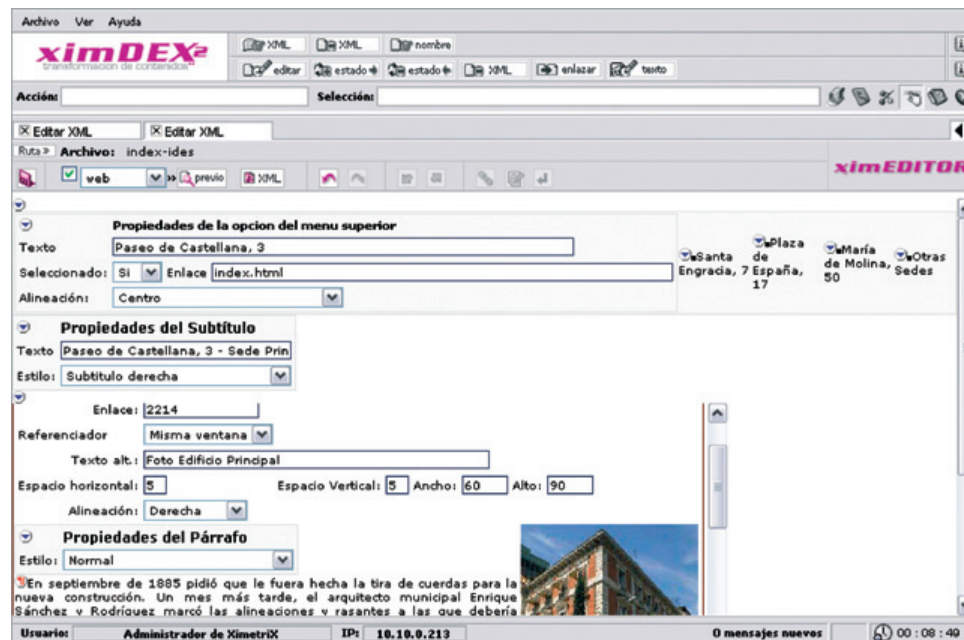


Figura 4: Edición de propiedades de los elementos de información.

El objetivo final del proceso de edición es la creación de un documento estructurado en un formato XML concreto (una noticia por ejemplo), documento interpretable por ximDEX para generar los contenidos y servicios en los formatos finales (xhtml, jsp, ...).

La figura siguiente recoge un documento XML tal y como es almacenado en el entorno:

```
<docxap idioma="es" nivel="noticias" documento=" 030428portalciudadano ">
<titulo>Noticias</titulo>
<noticia fecha="18-02-2004" imagen_asociada="" video_asociado="">
<titular> El Portal del Ciudadano gestionado con ximDEX por el MAP</titular>
<sumario>
<parrafo>El Portal del Ciudadano gestionado con ximDEX por el MAP</parrafo>
<parrafo>Miles de documentos, millones de enlaces</parrafo>
<parrafo>Gestión distribuida de contenidos</parrafo>
</sumario>
<parrafo>El 5 de Mayo de 2003 se presentó en Madrid la nueva versión del Portal del
Ciudadano, un proyecto en el que XimetriX ha participado para rediseñar su aspecto
grafico y su estructura de navegación y aplicar las ultimas tecnologías de gestión de la
información para la publicación automatizada de los contenidos, su adquisición y
gestión y su transformación a los múltiples formatos y canales soportados en el
proyecto.</parrafo>
<cuerpo>
<parrafo>Un elemento clave del éxito de la gestión de los contenidos es la plena
separación entre la información y su presentación, lo que permite su reutilización
automática por sistemas software. Un paso mas del MAP hacia la Web semántica de la
mano de XimetriX.</parrafo>
</cuerpo>
</noticia>
</docxap>
```

Figura 5: Ejemplo de Documento estructurado en ximDEX

Operando como un gestor de contenidos para la publicación de portales, los documentos estructurados, almacenados en el entorno, son convertidos automáticamente al formato final de publicación (XHTML, por ejemplo), mediante la aplicación recurrente de una serie de Plantillas de Transformación de Documentos (PTD), también descritas en XML.

Precisamente, gracias a la estructuración ligada al uso de anotaciones contextuales en forma de etiquetas XML, el mismo documento origen puede publicarse en su versión completa dinámica para Web y en una versión estática reducida para PDA.

Las plantillas de transformación son las encargadas de modificar cada elemento de información individual que forma el contenido o servicio en base a su contexto (sección que ocupa, perfil del usuario, relación con otros contenidos), canal (web, pda, móvil), idioma, y tecnología final de explotación (jsp, php, ...).

Para ello, las plantillas encapsulan las propiedades de presentación para cada plataforma y tienen pleno acceso al contenido y a su entorno (propiedades, meta-información, secciones), lo que permite al documento tomar decisiones dinámicas sobre su transformación (eliminar los párrafos menos relevantes de un documento para su visualización en una PDA; por ejemplo, el bloque de párrafos etiquetado como cuerpo en la figura anterior) o acceder a fuentes remotas de información (bbdd, web) para actualizar una información relevante que es generada dinámicamente.

3.2. Arquitectura de ximDEX

A diferencia de la mayoría de los gestores de contenidos existentes, donde se realiza una generación dinámica del documento publicado para cada visita al servidor de explotación, el gestor ximDEX desvincula por completo la creación y gestión de los documentos de su explotación.

Para ello, induce una separación nítida entre la etapa encargada de “la creación/importación, supervisión, transformación y publicación de los contenidos” y la etapa posterior de “explotación por servidores especializados” (de vídeo, de aplicaciones, web, etc.).

De este modo, el entorno ximDEX puede describirse como un “precursor” de contenidos y de sus componentes dinámicos, dejando la explotación a servidores especializados (ej.: apache, resin, weblogic) para cada tipo de contenido, lo que permite un mejor escalado de la infraestructura asociada y evita los “cuellos de botella” que inducen otros gestores de contenidos. Además, y lo que es más importante, permite adecuar la tecnología empleada a la explotación a realizar (PHP versus JAVA por ejemplo), lo que reduce el nivel de intrusividad de ximDEX en el ámbito de aplicación y aumenta las capacidades de interoperabilidad.

La Figura siguiente recoge el desglose de ximDEX en los módulos principales que lo componen y su vinculación con los servidores remotos encargados de la explotación:

El módulo dexC es el “orquestador” del sistema completo, encargándose de todos los eventos y procesos relacionados con la gestión de los contenidos almacenados y del envío de los documentos estructurados al módulo dexT, módulo encargado de la transformación, perfilado y personalización hasta llegar a los formatos finales a publicar en los servidores de explotación, tarea realizada por el módulo dexP de sincronización y de des-sincronización, mediante el acceso a los diferentes entornos de servidores (preproducción, producción, ...) definidos.

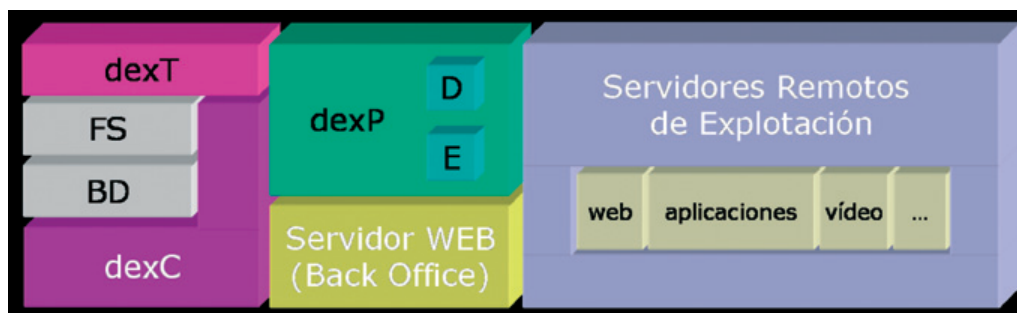


Figura 6: Módulos principales de ximDEX y servidores de explotación.

El módulo dexC garantiza además la integridad referencial de los contenidos y sus enlaces, la importación masiva de contenidos, la gestión de las anotaciones contextuales para labores de perfilado, los procesos de agregación de información desde fuentes remotas (bbdd, web), el control de versiones, el ciclo de vida de cada documento, etc.

3.3. Formatos libres y reutilización de contenidos

Un aspecto clave a garantizar en el proceso de Gestión de Contenidos es el libre acceso tanto a los documentos como a los formatos de intercambio y de transformación utilizados.

Este aspecto permite garantizar la reutilización futura de los contenidos, fomentando la interoperabilidad entre corporaciones mediante el intercambio de información en formatos no propietarios.

En el entorno ximDEX, la garantía de reutilización sencilla de los contenidos surge como consecuencia directa de las características siguientes del entorno:

- Permite la libre definición de los aplicativos XML utilizados por la corporación en la gestión de sus contenidos.
- Proporciona mecanismos potentes de estructuración y anotación de los mismos.
- Garantiza el pleno acceso a todos los contenidos y a los elementos que afectan a su transformación.
- Proporciona licencias libres a todos los sistemas encargados de realizar la transformación desde el documento estructurado XML a los formatos finales utilizados para la publicación y explotación.

4. EL ENTORNO XIMDEX EN EL PORTAL DEL CIUDADANO

XimetriX ha proporcionado la tecnología de gestión de contenidos ximDEX utilizada para la gestión de los elementos de información que configuran la versión de El Portal del Ciudadano actualmente publicada.

El Portal en su versión actual ha sufrido un cambio estético y de estructura para asentar las bases que permitieran la adición de nuevos idiomas y canales de publicación y el cumplimiento de requisitos de accesibilidad. Para ello, se ha abordado un ambicioso plan de reestructuración

del proceso de adquisición, gestión y publicación de la información que tiene como objetivos principales:

- Garantizar la calidad de la información almacenada y publicada.
- Garantizar su plena accesibilidad y la facilidad de migración a otros formatos electrónicos en el futuro.
- Facilitar la creación de elementos dinámicos de explotación por personal no técnico.

Para cubrir estos objetivos se ha reestructurado y unificado todo el modelo de datos, asociándole elementos descriptivos que faciliten su uso desde el gestor de contenidos ximDEX, lo que simplifica la creación de nuevos canales de publicación y permite la creación de productos interactivos de explotación de información por parte del personal del MAP.

De esta forma, la definición e implementación de un nuevo proceso de explotación de información ha pasado de implicar a múltiples departamentos y una petición externa de trabajo de varias semanas de duración, a gestionarse gráficamente por un único usuario de perfil redactor, lo que ha facilitado la definición de nuevos servicios en pocas horas.

El uso por el entorno ximDEX de formatos XML abiertos y escalables, tanto para describir los contenidos estructurados en el gestor como para describir las plantillas de transformación, es uno de los pilares para garantizar la reutilización presente y futura de los contenidos, facilitándose además el intercambio automatizado de información entre entidades en un marco de colaboración, lo que facilitará las tareas de agregación y sindicación entre instituciones y permitirá la manipulación automática de información en el paradigma ofrecido por la Web Semántica.

5. REFERENCIAS

- [Athe02] R. Athey, "Enterprise Content Management: taming content chaos", Search Results, Deloitte Consulting and Deloitte & Touche, ISBN 1-892384-21-7, 2002.
- [Bern01] T. Berners-Lee, J. Hendler et al, "The Semantic Web: A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities", Scientific American, vol. 284, no. 5, pp. 28-31, May 2001.
- [Pri04] J. A. Prieto, J. Villar et al, "Gestión Semántica de Contenidos con ximDEX", Proc. I Open Source World Conference, OSWC'04, Feb. 2004.
- [Zapt03] Zapthink LLC, "XML in the Content Lifecycle: Creating, Managing, Publishing, Distributing, Syndicating, and Protecting Content with XML", Jan. 2003.

BIBLIOGRAFÍA

- 1- <http://www.ximetrix.com/productos/ximDEX>