

# GESTIÓN SEMÁNTICA DE CONTENIDOS CON ximDEX

J. A. Prieto, J. Villar, D. Gómez, M. Estellés

*XimetriX network thoughts; Spain*

*ximlab@ximetrix.com*

La gestión de información y documentos mediante Sistemas de Gestión de Contenidos (CMS, *Content Management Systems*) ha evolucionado hacia la plena separación entre los contenidos y los elementos encargados de su presentación. Sin embargo, esta separación no está siendo aprovechada para dotar de significado a los elementos de información que constituyen los contenidos, lo que, en el paradigma de la Web semántica, facilitaría, por ejemplo, la sindicación de información sin necesidad de establecer manualmente las fuentes origen de la información y, a un nivel más avanzado, permitiría aumentar el grado de interoperatividad entre sistemas de distintas corporaciones.

En la actualidad, el uso de XML para la publicación en Web asocia, en la mayoría de los CMS existentes, una estructura y una sintaxis concreta, lo que dificulta considerablemente su posible uso para la realización automática de tareas de intercambio de información entre entidades en base a una sintaxis y/o semántica compartida. Por otro lado, los CMS tradicionales se constituyen además en los Servidores de Aplicaciones encargados de la explotación de los contenidos, lo que reduce considerablemente la interoperabilidad con otros sistemas, bloqueando además el acceso a los precursores de los contenidos.

El entorno ximDEX<sup>1</sup> de Gestión de Contenidos (CMS) desarrollado por XimetriX ([www.ximetrix.com](http://www.ximetrix.com)) proporciona los pilares necesarios para dotar de significado a los elementos de información que constituyen los contenidos. Para ello, proporciona a los usuarios la capacidad de edición y definición de los formatos finales a utilizar en la gestión de los contenidos (aplicativo XML), proporcionando además un motor de transformación con licencia libre (módulo dexT en ximDEX), encargado de la translación a los formatos finales de publicación (XHTML, HTML, PDF, ...) utilizados en los servidores Web y de explotación, a partir de los aplicativos XML libremente definidos. El entorno ximDEX proporciona además un editor de XML tipo *wysiwyg* vía web y simplifica o automatiza la mayoría de las tareas vinculadas al ciclo de vida documental, lo que garantiza un acceso sencillo a las tareas y facilita el uso del entorno por usuarios no avanzados.

La reutilización de contenidos ya existentes o heredados se garantiza mediante filtros de transformación automatizados. Sin embargo, en un *marco libre*, es aún más importante garantizar la reutilización futura de los contenidos y fomentar la interoperabilidad entre corporaciones en un entorno "colaborativo". En el entorno ximDEX, esta reutilización de los contenidos surge de forma natural como consecuencia directa de: i) permitir la libre definición de los aplicativos XML utilizados en la gestión de contenidos de una corporación y ii) estructurar y garantizar el pleno acceso a todos los contenidos y a los precursores de los mismos (documentos de origen, código fuente para explotación dinámica, plantillas de transformación de formatos, ...) mediante múltiples vías como, por ejemplo, Servicios Web.

Keywords: Gestión y Publicación de Contenidos, CMS, Web Semántica, Interoperabilidad.

---

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

## Contenidos y Web Semántica

La gestión de contenidos centrada en la publicación en Internet ha evolucionado gradualmente hacia la plena separación entre los contenidos y los elementos encargados de su presentación, es decir, de los aspectos visuales. El máximo exponente de esta separación lo proporciona el asentamiento de XML como tecnología para la publicación de contenidos [Zapt03, Athe02].

En el paradigma de la Web Semántica [Bern01] se asocia un significado a los elementos de información que forman los contenidos, proporcionándose simultáneamente mecanismos para estructurar y compartir ese significado en un dominio concreto de conocimiento. El dotar a los elementos de información que componen la Web actual de un significado bien definido permitiría, entre otros, mejorar las capacidades de búsqueda contextual, aumentar la interoperabilidad entre sistemas en entornos “de colaboración” y llegar, en combinación con Servicios Web, a “componer” de forma automática aplicaciones en base a servicios publicados [Shet02, Tsai03].

Sin embargo, en el marco de los Sistemas de Gestión de Contenidos (CMS, *Content Management Systems*) actuales, esta separación entre contenidos y presentación no está siendo aprovechada para añadir significado a los elementos de información que constituyen los contenidos, principalmente por la asociación realizada por los CMS existentes a una estructura y sintaxis predefinida y por la falta de madurez de las tecnologías de gestión de ontologías, técnica habitualmente usada para la representación de “conocimiento” en base a la especificación formal y explícita de términos en un dominio y de las relaciones existentes entre ellos [Chan99].

El acercamiento de los sistemas de Gestión de Contenidos a los paradigmas de la Web Semántica permitiría automatizar la realización automática de tareas de intercambio de información entre entidades en base a una sintaxis y/o semántica compartida, lo que facilitaría la sindicación y agregación automática de información, evitando la necesidad de establecer manualmente las fuentes origen de la información. A nivel más avanzado, permitiría aumentar el grado de interoperabilidad entre sistemas de distintas corporaciones, evolucionando de los actuales Portales de Contenidos a Portales de Servicios [Prie03].

*XimetriX network thoughts* ([www.ximetrix.com](http://www.ximetrix.com)) ha desarrollado ximDEX, un entorno para la adquisición, gestión y transformación de contenidos que permite la inclusión de significado bien definido a los elementos de información que forman los contenidos. El presente artículo recoge las principales características del entorno, resaltando los aspectos clave que posibilitan el tratamiento de contenidos en el paradigma que ofrecerá la Web Semántica.

### Gestión de Contenidos con ximDEX

El entorno ximDEX es un sistema automático de adquisición, transformación y publicación de contenidos en base a aplicativos XML sobre un entorno visual de “etiquetado” de los elementos de información. Así, en su configuración como CMS, el gestor ximDEX permite la automatización de la publicación de contenidos en formatos típicos de Internet, tanto de tipo estático como de tipo dinámico, y de todos los componentes encargados de su explotación (java, php, .net, etc.), a lo largo de todo el ciclo de vida del documento. El carácter del entorno facilita el escalado y desarrollo de nuevas aplicaciones, proporcionando simultáneamente mecanismos de reutilización de los contenidos y facilitando la asociación de significado a los elementos constitutivos de los contenidos.

El pilar central de esta operativa lo constituyen las capacidades de transformación de los documentos estructurados (documentos XAP –XML Active Publishers– en ximDEX), descritos en base a aplicativos XML libremente definidos por los usuarios, hacia los múltiples formatos finales de publicación e intercambio comúnmente aceptados en Internet.

Operando como CMS, los documentos XAP almacenados en el entorno son convertidos automáticamente al formato final de publicación (XHTML, por ejemplo), mediante la aplicación recurrentemente de una serie de Plantillas de Transformación de Documentos (PTD) que contienen las propiedades de presentación necesarias para cada uno de los canales vinculados a la publicación (pda's, web, móviles), respetando las características multilingüe requeridas. Los documentos no estructurados (PDFs, vídeos) son gestionados en base a metainformación aplicada mediante técnicas de anotación, controlando el entorno todo su ciclo de vida y versionado y la publicación y despublicación automática en los servidores vinculados al entorno de producción.

En el contexto de separación entre contenidos y presentación es necesario desvincularse de los entornos de producción existentes. Para ello, el entorno ximDEX gestiona los elementos de código encargados de los contenidos dinámicos, convirtiéndose en un precursor, tanto de los contenidos a publicar como de los elementos de explotación necesarios. Esto permite al entorno ximDEX operar de forma completamente separada respecto al entorno final de explotación, lo que disminuye considerablemente el nivel de intrusividad de la solución y aumenta el nivel de interoperabilidad.

### **Documentos estructurados y anotaciones contextuales**

Los documentos en ximDEX (documentos XAP) se estructuran mediante anotaciones XML elegidas libremente de un aplicativo XML predefinido por los usuarios del entorno, aplicativo con vinculación directa con las Plantillas PTD de transformación que encapsulan los aspectos necesarios para su traslación a otros sistemas de publicación, agregación y sindicación de información. Así, un formato XAP lo constituye el aplicativo XML, libremente definido por los usuarios del sistema, que proporciona las pautas para realizar la transformación adecuada mediante la invocación recurrente de las plantillas PTD.

Una vez incorporado (bien por agregación, bien por edición) el documento XAP al sistema, este es transformado de forma activa hasta llegar al formato final de publicación o intercambio requerido. Durante este proceso de transformación se expanden, de forma recurrente, las distintas anotaciones existentes en el documento, en base a un contexto predeterminado y acotado de publicación a elegir de una taxonomía de perfiles. Esta transformación implica habitualmente la inclusión de texto estructurado o no, permitiéndose también el acceso a fuentes remotas de información para la inclusión de información generada dinámicamente (*activa*) o incluso la ejecución de código interpretable.

En la línea de simplificación requerida para el acercamiento del entorno a personal no especializado, el entorno ximDEX proporciona un editor *wysiwyg* para XML, lo que permite la edición de los documentos estructurados sin conocer las etiquetas que lo componen, visualizando simultáneamente el estilo final del documento.

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

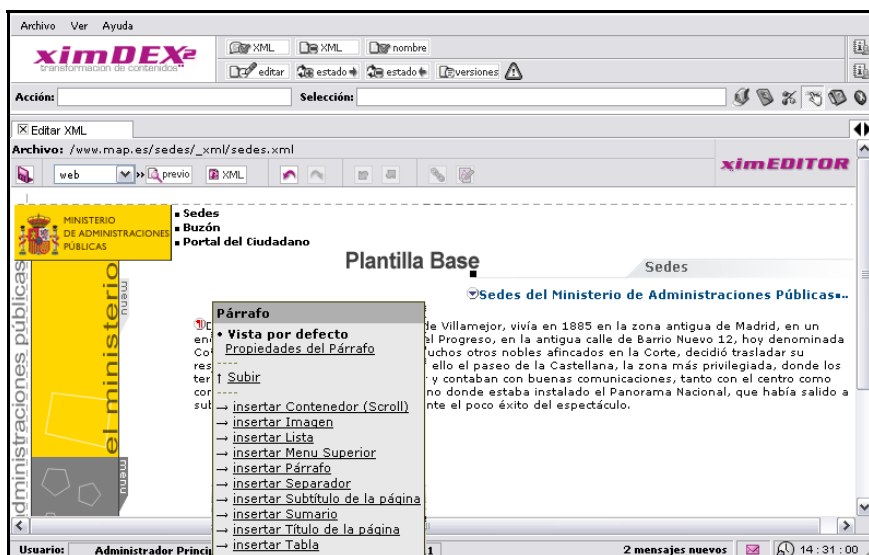


Figura 1: Edición *wysiwyg* de un documento XML en ximDEX.

## Módulos básicos de ximDEX

A diferencia de la mayoría de los gestores de publicación existentes, donde se realiza una generación dinámica de los contenidos publicados en cada visita al servidor de explotación, el gestor ximDEX desvincula la creación de los documentos de su explotación. Para ello, induce una separación 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) para cada tipo de contenido, lo que permite un mejor escalado de la infraestructura asociada y evita los “cuello de botella”, típicos de las arquitecturas multicapas con generación dinámica, inducidos por otros CMS. 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:

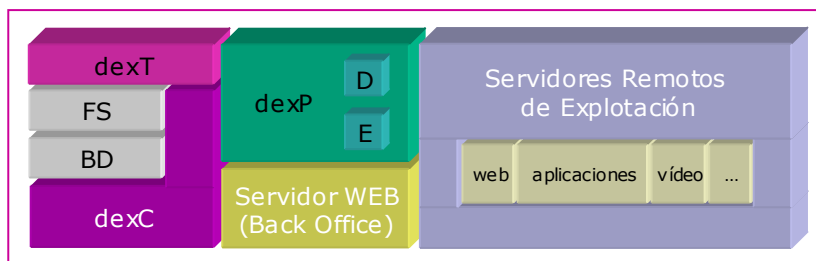


Figura 2: Módulos principales de ximDEX y servidores de explotación.

El módulo dexC es el “orquestador” del sistema completo. Para ello, accede a los datos y documentos, estructurados (XML) o no (HTML, PDF, etc.), almacenados en la base de datos del sistema (BD), invoca al módulo de transformación dexT, encargado de realizar la síntesis real de los contenidos estructurados en los diversos idiomas y formatos, gestiona el ciclo de vida completo de cada documento y sus ventanas de publicación y despublicación en los

formatos finales para sincronización a los servidores remotos, agrega información de fuentes externas y realiza importaciones masivas de contenido aplicando filtros de transformación, lo que facilita la reutilización de contenidos heredados, realiza un control de la integridad de los enlaces internos y externos, la gestión de las anotaciones contextuales, etc.

El módulo dexT constituye el corazón de ximDEX. Se encarga de generar las diversas “variantes” de los elementos de un documento a partir del documento original descrito en formato XAP. Los documentos y elementos generados por el módulo dexT de ximDEX se encuentran ya en el formato final de presentación (HTML, código PHP o clases java, por ejemplo), para su sincronización final al servidor elegido mediante el módulo dexP. El dexT se encarga además de controlar las diversas “variantes” de un documento, es decir, los diversos idiomas en que se publica, las versiones para impresión, y para los diversos canales de publicación (PDA, móvil, Web). El formato XAP es el aplicativo XML que proporciona las pautas para realizar la transformación adecuada mediante la invocación recurrente de las plantillas PTD definidas.

Por último, el módulo dexP se encarga de enlazar el entorno ximDEX con los servidores remotos de explotación que publican y sirven los documentos en los formatos finales (servidores de aplicación, vídeo, web, ...). Este módulo presenta dos personalidades entrelazadas, una de tipo estático y otra de tipo dinámico, realizando la sincronización con el servidor de explotación adecuado en la fecha y hora indicada por el perfil de publicación, o bien, en caso de optarse por una explotación dinámica del contenido directamente por el entorno ximDEX, realiza la petición al módulo de transformación (dexT) para la interpretación del documento XAP a petición del servidor Web de explotación.

La figura siguiente recoge una de las pantallas de gestión de flujos de trabajo que controlan el comportamiento del módulo dexC:

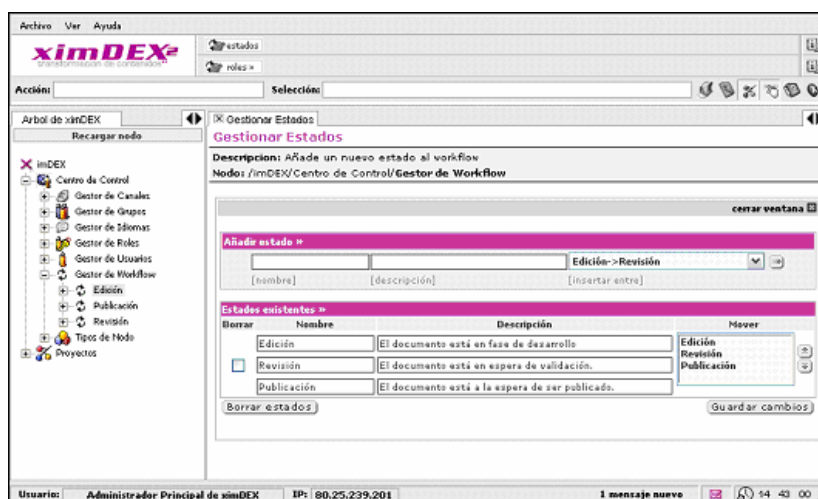


Figura 3: Control de Flujos de Trabajo en ximDEX.

## Formatos libres y reutilización de contenidos

Un aspecto clave, habitualmente olvidado de forma intencionada por otros entornos y productos, es el libre acceso a los documentos y contenidos almacenados en el gestor y la publicación de los formatos de intercambio utilizados [Cosm97]. En un marco libre es aún más importante garantizar la reutilización futura de los contenidos y fomentar 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 los aspectos siguientes relacionados con la información y los formatos: i) permitir la libre definición de los aplicativos XML (formato XAP) utilizados por la entidad o corporación en la gestión de *sus* contenidos, ii) proporcionar mecanismos potentes de estructuración y anotación de los mismos, y iii) garantizar el pleno acceso a todos los contenidos (XAP y generados) y a los precursores de los mismos (código fuente para explotación dinámica, plantillas de transformación PTD, ...), a través de múltiples vías como, por ejemplo, Servicios Web.

Otro aspecto fundamental es proporcionar con licencias libres todos los sistemas encargados de realizar la transformación de los contenidos desde el formato original (XAP) a los formatos finales utilizados para la publicación. En el entorno ximDEX, esta transformación la realiza el módulo dexT, publicado con licencia libre, capaz de transformar y estructurar portales complejos de contenidos multilingüe a múltiples formatos finales.

## Conclusiones

La separación creciente entre contenidos y presentación abre un camino directo hacia la inclusión de técnicas provenientes de la Web Semántica, lo que permitirá asociar significado concreto y unívoco, en un dominio de conocimiento, a los elementos de información que forman los contenidos.

En este contexto, se han presentado las características fundamentales del gestor ximDEX desarrollado por XimetriX ([www.ximetrix.com](http://www.ximetrix.com)), entorno que facilita la asociación de semántica a los elementos de información durante la creación de documentos, resaltándose además los requisitos fundamentales para garantizar la reutilización futura de los contenidos y el intercambio automatizado de información entre entidades en un marco de colaboración, lo que facilitará las tareas de agregación y sindicación y permitirá la manipulación automática de información en el paradigma ofrecido por la Web Semántica.

## 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.
- [Chan99] B. Chandrasekaran, V. R. Benjamins et al, "What Are Ontologies, and Why Do We Need Them?", IEEE Intelligent Systems, pp. 20-26, Jan./Feb. 1999.
- [Cosm97] P. Di Cosmo, "Piège dans le Cyberspace / Trampa en el cyberspacio / CyberSnare", <http://www.pps.jussieu.fr/~dicosmo/Piege/PiegeES.html> (español)
- [Prie03] J. A. Prieto, "Interoperability Dossier", Internal Report, Jul. 2003.
- [Shet02] A. Sheth, C. Bertram et al, "Managing Semantic Content for the Web", IEEE Internet Computing, Volume: 6, Issue: 4, pp. 80-87, Jul-Aug. 2002.
- [Tsai03] Tse-Ming Tsai, Han-Kuan Yu et al, "Ontology-Mediated Integration of Intranet Web Services", Computer Magazine, Volume:36, Issue:10, pp. 63-71, Oct. 2003.
- [Zapt03] Zapthink LLC, "XML in the Content Lifecycle: Creating, Managing, Publishing, Distributing, Syndicating, and Protecting Content with XML", Jan. 2003.