

MP_0373. Lenguajes de marcas y sistemas de gestión de información

UF5. Almacenamiento y manipulación de la información

5.2. Herramientas de tratamiento y almacenamiento de información



Índice

Objetivos	3
Introducción	4
Herramientas	5
Herramientas de creación de documentos XML	5
Herramientas de bases de datos	6
Bases de datos XML-Enabled	6
Bases de datos XML Nativas	7
Herramientas de BD XML nativas	7
Tamino	7
EXist	8
BaseX	8
Sedna	9
Otras herramientas	9
Despedida	11
Resumen	11

Objetivos

Con esta unidad perseguimos los siguientes objetivos:

- Conocer las **herramientas** actuales de manejo de XML.
- Conocer algunas herramientas de **creación de documentos XML**.
- Conocer las herramientas de **bases de datos XML**.
- Conocer las herramientas actuales de **bases de datos XML-Enabled**.
- Conocer las herramientas actuales de bases de datos **XML Nativas**.

Introducción

Las herramientas para manipular XML se pueden dividir en dos grupos:

- Herramientas de **creación**.
- Herramientas de **administración de contenido**.

Muchas de las bases de datos del mercado trabajan con documentos XML, que insertan como datos en la base de datos relacional, sin embargo existen otros productos más modernos que se han creado bajo el ámbito de las bases de datos XML nativas, e incluso se han diseñado para poder incorporar estándares de transformación de XML y de consultas XML. En este capítulo revisamos el estado de todas ellas.

Las herramientas que existen en el mercado en la actualidad para trabajar con XML se pueden dividir en dos grandes grupos:

Herramientas de creación de documentos XML

Herramientas que nos permiten construir XML válidos usando DTD, XML Schema, Schematron etc. Solo servirían para incluir datos en un conjunto de documentos XML.

Herramientas de bases de datos

Serían las que organizarían conjuntos de XML que formarían las bases de datos que hemos estudiado con anterioridad. En este apartado se podrían incluir las bases de datos relacionales XML-Enabled y las nativas.

Herramientas

Herramientas de creación de documentos XML

Las herramientas de creación de documentos XML deben tener las siguientes características:

- **Crear documentos** desde cero o en base a DTD, o XML Schemas.
- **Editar documentos XML**, pudiendo insertar, manipular y borrar datos.
- Editar **de forma textual o visual**.
- **Validar documentos** en base a DTD, XML Schema, Schematron o Relax NG.
- Comprobar **espacios de nombres**.
- **Realizar transformaciones en base a XSLT**.
- Procesar datos con el modelo **XPath**.
- **Creación de DTD**.
- **Creación de XML Schema**.
- **Explorador** de esquemas XML.
- **Esquematzación de documentos**, así los elementos se pueden expandir y contraer.
- Capacidad de **generar esquemas** de lenguaje de definición de esquemas XML (XSD) a partir del documento de instancia XML.
- Creación de **estilos XSL**.

Entre las más importantes destacan:

- **SoftQuad Xmetal**. Es una aplicación de software utilizada para crear y editar documentos en XML y SGML. Necesita que los documentos XML estén validados.
- **Adobe FrameMaker+SGML**. Proporciona una GUI completa que manipula documentos SGML/XML. Permite abrir documentos no válidos y rectificarlos.
- **IBM Xena**. Herramienta desarrollada en Java. Proporciona una GUI para editar documentos válidos.
- **Bluestone Visual-XML**. Herramienta de creación XML orientada a la creación de soluciones de bases de datos empresariales que incluyen XML. Posee un asistente para la publicación de bases de datos que genera DTD y plantillas de documentos XML a partir de tablas de bases de datos. Puede descargarse gratuitamente en: <http://www.bluestone.com/xml/visual-XML>.
- **XML Notepad**. Se trata de una herramienta poco potente de Microsoft que sirve para realizar tareas de creación sencillas. Para validar documentos requiere tener instalado IE5.0. Puede descargarse gratuitamente en <http://msdn.microsoft.com/xml/notepad>.
- **XMLSPY**. Editor de XML y entorno de desarrollo integrado. Este software desarrollado por Altiva permite a los desarrolladores crear aplicaciones basadas en XML y servicios web utilizando tecnologías como XML, XML Schema, XSLT, XPath, XQuery entre otras. <http://www.xmlspy.com/>.
- **XMLwriter**. Proporciona funcionalidades de forma sencilla y eficiente para la realización de tareas de formato, validación, transformación y estilo. <http://xmlwriter.net/>.

Herramientas de bases de datos

El software que nos permite crear datos basados en documentos XML e implementa los módulos de manejo y búsqueda de datos se clasifica por dos tipos:

- Software de **base de datos XML-Enabled**, que son las bases de datos clásicas que se han adaptado a trabajar con XML.
- Software que implementa **bases de datos XML nativas**.

Bases de datos XML-Enabled

Las **bases de datos XML-Enabled** son aquellas que incorporan la información de un documento XML en su **correspondiente esquema relacional o de objetos**, por lo que contienen funcionalidad para poder manejar la transferencia de datos entre documentos XML y las estructuras de datos propias.

Son utilizadas por las aplicaciones centradas en datos, excepto cuando la base de datos soporta almacenamiento XML nativo.

La diferencia fundamental entre estas últimas es que las XML-Enabled utilizan estructuras específicas que deben ser mapeadas al documento XML en tiempo de diseño.

Podemos decir también que un **SGBD XML-Enabled solo puede manejar y almacenar los documentos que encajan dentro del modelo definido para ellos**, mientras que un SGBD XML nativo debe manejar todos los tipos de documentos posibles.



IBM DB2 es un sistema de base de datos de nivel de entrada diseñado para el proceso de transacciones y la gestión de cargas de trabajo de consultas complejas. Proporciona funciones a escala de empresa y está optimizada para utilizar hasta 8 núcleos de procesador y 8 GB de memoria.



Oracle es un gestor de base de datos relacional que hace uso de los recursos del sistema informático en todas las arquitecturas de hardware para garantizar su aprovechamiento al máximo en ambientes cargados de información. Es el conjunto de datos que proporciona la capacidad de almacenar y acudir a estos de forma recurrente con un modelo definido como relacional. Además es una suite de productos que ofrece una gran variedad de herramientas. Es el mayor y más usado Sistema Gestor de Base de Datos Relacional (RDBMS) en el mundo. La Corporación Oracle ofrece este RDBMS como un producto incorporado a la línea de producción. Además incluye cuatro generaciones de desarrollo de aplicación, herramientas de reportes y utilitarios.



PostgreSQL es un SGBD relacional orientado a objetos y libre, publicado bajo la licencia BSD.

Como muchos otros proyectos de código abierto, el desarrollo de PostgreSQL no es manejado por una empresa y/o persona, sino que es dirigido por una comunidad de desarrolladores que trabajan de forma desinteresada, altruista, libre y/o apoyada por organizaciones comerciales. Dicha comunidad es denominada el PGDG (PostgreSQL Global Development Group).

Bases de datos XML Nativas

Una base de datos XML Nativa es aquella que:

- **Define un modelo de datos XML.** Define qué elementos son lógicamente significativos. Todos tendrán en cuenta los elementos, atributos, texto y orden en el documento, aunque es posible que de un sistema a otro los elementos reflejados en el modelo varíen.
- Utiliza el documento como unidad mínima de almacenamiento.
- **Puede usar cualquier estrategia de almacenamiento.** El almacenamiento físico de los documentos puede realizarse sobre un SGBD relacional tradicional, sobre ficheros con estructura propia o cualquier otro método.

La principal diferencia es que **una base de datos XML Nativa provee de su propio modelo de datos**, mientras que un sistema XML-Enabled tiene su propio modelo de datos y añade una capa de software que permite de alguna manera almacenar documentos XML y recuperar los datos generando nuevos documentos XML.

Herramientas de BD XML nativas

Algunas bases de datos XML Nativas son:

Tamino

Tamino es el SGBD nativo de la empresa SoftwareAG. Es un producto comercial de alto rendimiento y disponibilidad, además de ser uno de los primeros SGBD XML nativos disponibles.

Native XML Data Store + XML Engine es el componente central de la arquitectura, contiene el parser XML, el almacén nativo de datos y el intérprete de consultas.

- **Arquitectura**
 - **Data Map.** Almacén de metadatos, contiene información sobre cómo validar esquemas, almacenar e indexar información, mapeo de estructuras, etc.
 - **X-Node.** Es el componente para dar acceso a bases de datos externas, mapeando los datos a estructuras XML. El acceso a esta información es transparente para el usuario y se accede a ella cada vez que se necesita, es decir, que no es replicada.

- **X-Tension.** Permite la definición de funciones de usuario para ampliar las prestadas por Tamino.
- **Tamino Manager.** Herramienta gráfica de gestión.
- **Almacenamiento de documentos.** Los documentos se almacenan en una base de datos propia y no se transforman en otro modelo. Existe un espacio separado para documentos y para índices.

Un *doctype* es el elemento raíz de un DTD o XML Schema, es decir, el elemento que define el comienzo y el final del documento.

La base de datos está estructurada en colecciones. Una colección es un conjunto de documentos, de modo que es una estructura de árbol donde cada documento pertenece a una única colección.

EXist

EXist-db es un sistema de gestión de bases de datos libre y de código abierto que almacena datos XML de acuerdo a un modelo de datos XML.

Algunas de sus características son: soporte para distintos lenguajes de consultas XML como XQuery, XPath o XSLT, indexación de documentos y soporte para la actualización de los datos y para multitud de protocolos como SOAP, XML-RPC, WebDav y REST. En la actualidad cumple el estándar XQuery en un 99.4%.

Descarga y documentación

Para descargar la aplicación y examinar la documentación puedes visitar este enlace.

<http://exist-db.org/exist/apps/homepage/index.html>

BaseX

BaseX es una base de datos XML escalable, con un alto rendimiento y muy ligera. Con procesador XPath/XQuery y soporte de la última actualización del W3C, es una opción idónea para nuestro trabajo.

Entre sus características se encuentran:

- Alto rendimiento de almacenamiento con texto, atributos, texto completo y rutas.
- Apoyo de las recomendaciones del W3C XPath/XQuery.
- Una de las más altas tasas de cumplimiento para todas las especificaciones admitidas.
- Arquitectura cliente/servidor con soporte de transacciones seguras ACID, gestión de usuarios y autenticación.
- Visualizaciones altamente interactivas, con soporte para documentos XML de gran extensión.
- Editor XQuery en tiempo real disponible, con resaltado de sintaxis y *feedback* de errores.
- Amplia gama de interfaces: REST/RESTXQ, WebDAV, XQJ, XML:DB.

Descarga

Puedes descargar el software en este enlace. <http://basex.org/download/>

Sedna

Sedna es un sistema de gestión de base de datos XML nativa de código abierto, implementado a partir de cero con todas las funciones de administración de una base de datos para almacenar grandes cantidades de datos XML.

La aportación especial de la base de datos Sedna es su estrategia de almacenamiento mejorado basado en esquemas (*schema-driven storage strategy*), que permite gran eficiencia a la hora de procesar las consultas y almacenamientos en la base de datos. Tiene un sistema de manejo de memoria muy novedoso.

Dentro del apartado de almacenamiento debemos mencionar también un elemento novedoso de Sedna: las relaciones de inclusión entre elementos; se utilizan punteros directos para representar relaciones de un nodo XML, como padre-hijo con sus respectivos hermanos. Este método, a diferencia de enfoques relacionales que requieren realizar "joins" para atravesar un documento XML, permite realizar esto a través de un puntero directo.

Descarga y documentación

Puedes descargar el software en este enlace. <https://www.sedna.org/>

Otras herramientas

- **Qexo:** escrito en Java y con licencia GPL, se distribuye integrado dentro del paquete Kawa; un IDE (Entorno de Desarrollo Integrado) que permite crear aplicaciones en lenguaje Java.
- **Saxon:** escrito en Java y distribuido en dos paquetes:
 - **Saxon-B** es *open-source* bajo licencia GPL y contiene una implementación básica de XSLT 2.0 y XQuery.
 - **Saxon-SA**, contiene un procesador completo XSLT y XQuery, pero tiene licencia propietaria, aunque está disponible una licencia de evaluación de 30 días.
- **Qizx/open:** es una implementación Java de todas las características del lenguaje, excepto la importación y validación de XML-Schemas. Es uno de los motores con licencia GPL más completo que existe.
- **Xquark Bridge:** es una herramienta que permite importar y exportar datos a bases de datos relacionales utilizando XML, ofreciendo, por tanto, la posibilidad de manejar estructuras XML y realizar la transformación a objetos de la base de datos y viceversa. Además, XQuark respeta las restricciones de integridad y transforma las relaciones implícitas en los documentos XML en relaciones explícitas en la base de datos. También soporta la consulta y manipulación de los datos en formato XML utilizando el lenguaje XQuery. XQuark-Bridge soporta MySQL, Oracle, SQLServer y Sybase y tiene una licencia LGPL.

- **BumbleBee:** es un entorno de prueba automático, creado para evaluar motores de XQuery y validar consultas XQuery. BumbleBee permite, entre otras cosas, comprobar si una versión más moderna de nuestro motor permite seguir ejecutando nuestras consultas. BumbleBee se distribuye con un conjunto de pruebas ya preparadas y además ofrece un entorno sencillo para redactar y ejecutar nuestras propias pruebas.

Despedida

Resumen

Has terminado la lección, veamos los puntos más importantes que hemos tratado.

En esta unidad hemos conocido algunas de las herramientas con las que podemos trabajar con los datos XML, tanto para su creación como para la gestión de bases de datos.

- Las herramientas de creación nos permiten construir XML válidos usando DTD, XML Schema, Schematron etc. Estas solo servirían para incluir datos en un conjunto de documentos XML.
- Las herramientas de bases de datos son las que organizarían conjuntos de XML que formarían las bases de datos que hemos estudiado con anterioridad. Éstas pueden ser bases de datos relacionales XML-Enabled o nativas.