

Recuperación y organización de la información

Bases de datos nativas en Internet y sistemas de almacenamiento y recuperación de documentos HTML, RDF y XML

1. Introducción

Bases de datos

Las **bases de datos** son, en su acepción más sencilla, una reunión de datos almacenados para poder ser recuperados cuando corresponda. Estos datos están organizados y se relacionan entre sí.

Normalmente, se suele denominar **base de datos** a aquel conjunto de datos que se encuentran almacenados en un soporte informático.

Existen diferentes tipos de **bases de datos**, pero nos vamos a centrar en las bases de datos nativas, de las que se ocupa este trabajo

Sistemas de recuperación y almacenamiento de documentos

Estos sistemas permiten almacenar documentos por un período de tiempo y permitir el acceso a dichos materiales cuando éstos sean necesarios, mediante búsquedas específicas.

Entre estos sistemas, existen algunos dedicados a la recuperación de ciertos tipos de documentos en formato electrónico.

2. Bases de datos nativas

La necesidad de procesar y almacenar XML ha hecho que salgan a la luz diferentes tipos de herramientas de software. Una de ellas han sido las bases de datos nativas.

Una base de datos nativa define modelos lógicos para documentos XML, y almacena y recupera documentos acordes a ese modelo. Como mínimo, el modelo debe incluir elementos, atributos, PC DATA y el orden del documento.

Estas bases de datos, como cualquier otra, permiten transacciones, que accedan varios usuarios a la vez, etc. lo único que cambia es el formato que soportan, ya que, como hemos dicho, están especializadas en documentos XML.

Este tipo de bases de datos están especializadas en almacenar documentos y datos XML y almacenar todos los componentes del modelo XML intactos.

Las bases de datos nativas XML no representan un nuevo modelo de bases de datos, y no van a suplantar a las bases de datos existentes. Son una herramienta que ofrece un buen almacenamiento y manipulación de los documentos XML.

Las **características** principales de estas bases de datos son:

- Emplean como unidad lógica fundamental de almacenamiento el documento XML.
- Define un modelo (lógico) para un documento XML.
- Almacena y recupera documentos de acuerdo a ese modelo.
- Soportar lenguajes de consulta XML.
- Preservar el orden del documento, las instrucciones de procesamiento, los comentarios, las secciones CDATA y las entidades.
- No se requiere un modelo subyacente de almacenamiento.
- Puede ser una base de datos relacional, orientada a objetos o jerárquica.

3. Sistemas de almacenamiento y recuperación

¿Qué es un sistema de recuperación y almacenamiento?

Este tipo de sistemas se encargan de mantener una serie de datos para, posteriormente, poder acceder a ellos sin dificultad.

Existen diversos tipos de sistemas de almacenamiento y recuperación, pero aquí vamos a poder observar algunos ejemplos orientados a documentos RDF, HTML y XML.

Ejemplos de sistemas de almacenamiento y recuperación de documentos HTML, RDF y XML

Sesame

Sesame es un marco Java de fuente abierta para almacenar, hacer consultas y razonar con RDF y RDF Schema.

Puede utilizarse como base de datos para RDF y RDF Schema, o como una biblioteca Java para aplicaciones que necesitan trabajar internamente con RDF.

Por ejemplo, en el supuesto en el que se necesita leer un gran archivo RDF, encontrar la información relevante para una aplicación y usar esa información. Sesame ofrece las herramientas necesarias para analizar esa información, interpretarla, hacer consultas y almacenarla, embebida en la propia aplicación o, si se prefiere, en una base de datos separada o incluso en un servidor remoto.

En resumen, con Sesame tenemos disponible una serie de herramientas para gestionar RDF fácilmente.

Jena

Es un marco que ha sido desarrollado por HP Labs Semantic Web Programme. Se utiliza para manipular metadatos desde una aplicación Java.

Existen dos versiones:

JENA 1

- Principalmente soporte para RDF.
- Capacidades de razonamiento limitadas.

JENA 2

- Incluye además una API para el manejo de Ontologías.
- Soporta el lenguaje OWL.

Opera XML Store

Opera XML Store es una solución de almacenamiento para datos XML. El servicio está pensado para desarrolladores de aplicaciones que necesitan almacenamiento de datos disponible en diferentes plataformas.

Podría usarse, por ejemplo, para mantener una aplicación en el teléfono móvil y en el ordenador de manera sincronizada

Los datos que se han almacenado pueden ser recuperados en dos formatos diferentes, XML o HTML.

4. Tipos de formato

HTML

HTML es el lenguaje con el que se escriben las páginas web. Podemos decir por lo tanto que el HTML es el lenguaje usado por los navegadores para mostrar las páginas web al usuario, siendo hoy en día la interfaz más extendida en la red.

Este lenguaje nos permite aglutinar textos, sonidos e imágenes y combinarlos a nuestro gusto.

Además, el HTML nos permite la introducción de referencias a otras páginas por medio de los enlaces hipertexto.

XML

XML es un Lenguaje de Etiquetad de Extensible muy simple, pero estricto, que juega un papel fundamental en el intercambio de una gran variedad de datos.

Es un lenguaje muy similar a HTML pero su función principal es describir datos y no mostrarlos como es el caso de HTML.

XML es un formato que permite la lectura de datos a través de diferentes aplicaciones.

Las tecnologías XML son un conjunto de módulos que ofrecen servicios útiles a las demandas más frecuentes por parte de los usuarios.

XML sirve para estructurar, almacenar e intercambiar información.

RDF

RDF, cuyas siglas significan Resource Description Framework, en castellano, Marco para la Descripción de Recursos, es utilizado para el procesamiento de metadatos.

Con RDF, pueden interoperar distintas aplicaciones, intercambiando información que puede leerse a través del ordenador en la Red.

Asimismo, con RDF es fácil habilitar el procesamiento automatizado de recursos Web.

El modelo de datos básico de RDF contiene tres tipos de objetos: recursos, propiedades y sentencias.