

## Síntesis conceptual

<b>Asignatura:</b> Acceso a datos
<b>Unidad:</b> 6. XML

### Resumen

XML (Extensible Markup Language) es un lenguaje de marcas estandarizado por la W3C. Una de sus principales características es la compatibilidad con la gran mayoría de sistemas, por eso su uso es muy extendido.

La estructura de XML está compuesta por nodos representando una jerarquía, el nodo superior se le denomina raíz o nodo padre. Siempre hay que indicar el encabezado de especificación `<?xml version="1.0" encoding="utf-8"?>` como primera etiqueta. Cada elemento está formado por el nombre de la etiqueta y si tiene atributos, nombre del atributo, igual y su valor.

Para verificar que un documento XML está bien formado, existen técnicas para validar:

- Mediante DTD. Solo se centran en la estructura de los elementos, su orden y número de apariciones.
- Mediante Esquemas XML. Existe un esquema con la especificación de cómo tiene que estar formado el documento XML, denominado XSD. Es más fiable que DTD.

En el documento XML tiene que especificar en su cabecera que técnica de validación usará y el documento.

La gran mayoría de lenguajes de programación puede trabajar directamente con XML, para ello deben de poseer parsers o analizadores que compruebe la sintaxis y una vez verificado pueden acceder al contenido del documento XML. Una de las técnicas de parsing es emplear el DOM, que permite tener en memoria todos los elementos y sus valores en una estructura arbórea. Una de las desventajas que tiene DOM es que no es eficiente para documentos XML que tienen una cantidad muy grande de información, debido a que necesita cargar todo en memoria. SAX es otro parser que permite hacer lo mismo que DOM pero a través de eventos. No necesita cargar todo el documento en memoria, si no, trae a memoria aquella parte necesitada según se produce un evento.

A parte de los parser, XML posee un lenguaje de consulta denominado Xpath. Es posible acceder a cualquier elemento mediante expresiones en Xpath. Las expresiones empiezan usando `/` e indicando el nombre del nodo raíz separado por tantas `/` y nombres de subniveles al que se quiera acceder.

Otra de las posibilidades que se puede usar con XML es traducir su contenido en una colección de objetos a partir de la API JAXB de java, a este proceso se denomina Binding. Para ello, es necesario disponer de un compilador de binding intermedio que se encarga de generar las clases necesarias según un esquema XML.

## Conceptos fundamentales

- **DOM:** (Document object model) Los documentos xml se representan en su estructura arbórea permitiendo navegar a cualquier nodo, acceder, modificar, añadir o quitar elementos.
- **SAX:** (Simple API for XML) Es un API estándar para JAVA que permite analizar o parsear documentos XML basado en eventos.
- **XSD:** Fichero con la especificación del esquema de validación XML
- **Serialización:** Proceso de transformar una estructura de datos no secuencial en un fichero de datos secuencial.
- **Deserialización:** Proceso de transformar un fichero de datos secuencial en una estructura de datos no secuencial