



Tarea: trabajo con XML en Java

Alberto Mateos Checa

Implementación de dos formas de trabajo con XML en Java, mediante DOM y SAX

1. Introducción.-

XML es un formato estándar de intercambio de información y datos, por lo que es posible trabajar con este formato en Java. Además, existen diversas formas de tratar este tipo de documentos. En particular, nos vamos a centrar en DOM y SAX ya que son dos de las formas de trabajo con XML más comunes aunque, como veremos, ambas son radicalmente distintas.

2. DOM.-

DOM (Document Object Model) es una forma de trabajo con XML en la que se accede a cada uno de los niveles del documento formando un árbol. De esta forma, se puede ir navegando a través de las diferentes etiquetas hasta llegar a obtener o almacenar la información deseada.

Así, DOM se presenta como una forma de tratar los XML bastante rápida y útil en casos en los que el documento sea de tamaño reducido o bien se conozca a la perfección la estructura del XML, pudiendo accederse directamente a la información buscada en caso de conocer dónde se encuentra de forma exacta. En caso contrario, si no se conoce la estructura del XML, puede ser una forma de trabajo tediosa y compleja ya que habría que ir descubriendo qué se almacena en el XML.

Para el uso de DOM en Java se hace uso de los paquetes `javax.xml` y `org.w3c.dom` que permite trabajar con documentos XML y con DOM respectivamente. De este modo, para trabajar con estas tecnologías hay que crear un parser haciendo uso de las clases `DocumentBuilderFactory` y `DocumentBuilder`. Una vez creado el parser, se le pasa el documento XML.

3. SAX.-

SAX (Simple API for XML) es otra forma de trabajo con documentos XML cuya orientación es completamente diferente a la usada en DOM. En este caso lo que se pretende es analizar el XML en busca de información de una forma más rápida y eficiente.

El funcionamiento de SAX es sencillo: se recorren de forma automática todas las etiquetas que componen el documento XML y cuando se encuentra una etiqueta SAX que nos interesa se realiza alguna acción. De esta forma, este método se presenta como más rápido y útil en caso de tener documentos extensos o documentos en los que no se conoce, a priori, la estructura.

Para el uso de SAX en Java hay que importar los `javax.xml` y `org.xml.sax`. De esta manera, de igual modo que con DOM, hay que crear primero un parser al que después se le pasará el documento XML.

Posteriormente, se recorre el documento XML, ejecutándose ciertos métodos predefinidos, cuando se produzcan, eventos que habrá que sobrescribir para que realicen la función que queramos. En este caso, los métodos que nos interesan son **startElement** y **endElement**. El primero de ellos se ejecuta cuando se encuentra una etiqueta de inicio en el XML e incluye ciertos parámetros que se le pasan como, por ejemplo, el nombre de la etiqueta facilitando al programador reconocer la etiqueta analizada y realizar una acción o no según se requiera. El segundo de los métodos se ejecuta cuando el parser encuentra una etiqueta de fin de un elemento y se comporta de igual manera que el primero.

4. XML en otros lenguajes.-

A pesar de que el tratamiento de documentos XML en Java es relativamente poco complejo, hay que decir que existen otros lenguajes en los que trabajar con archivos XML es mucho más sencillo. Este es el caso de lenguajes de scripting rápido como Python, Perl o Ruby y otros lenguajes destinados a su uso en Internet como Javascript o PHP.

De esta forma, los lenguajes citados anteriormente suelen disponer de librerías que hacen mucho más sencillo el trabajo con documentos XML. En la entrega de esta tarea **se incluyen algunos ejemplos de tratamiento de documentos XML en Ruby y Python utilizando tanto DOM como SAX.**