

Procesamiento de Formatos en Aplicaciones Telemáticas: XML y DTDs

En esta unidad se introducen los conceptos básicos de XML (*eXtensible Markup Language*) así como su validación mediante DTDs (*Document Type Definitions*). Se trabajará en ello en una sesión de teoría y una sesión de prácticas.

Referencias básicas

Se recomienda consultar el libro *Learning XML, Second Edition*, Erik T. Ray, O'Reilly Media Inc. (2003), disponible electrónicamente a través de la Biblioteca en el sistema *Safari Books Online*. El libro está también accesible desde fuera de la red de la Universidad si el usuario se conecta a la Web de la Biblioteca y se autentica con su usuario y contraseña de UC3M.

Más concretamente, los siguientes fragmentos del libro están relacionados con la materia de esta unidad:

- Conceptos básicos y estructura de un documento XML: capítulo 2 (*Markup and Core Concepts*).
- DTDs: apartado 4.2 (*DTDs*).

Prácticamente cualquier otro libro acerca de XML que se encuentre disponible será igualmente útil para esta parte de la materia.

A modo de referencia, se remite al alumno a la propia [especificación de XML por parte del W3C](#). Es el documento que especifica el estándar de XML, pero puede resultar útil para realizar consultas acerca de aspectos puntuales del lenguaje, así como referencia última en caso de dudas.

Por otra parte, en el sitio Web [w3schools](#) se pueden encontrar buenos manuales de [XML](#) y de [DTD](#).

Sesión 1: grupo magistral

En esta sesión se realiza una introducción a los aspectos más básicos de XML (estructura de un documento XML, sistemas de codificación de caracteres, aplicaciones de XML, concepto de documento bien formado), así como a los DTDs y el concepto de documento XML válido.

Sesión 2: práctica

El objetivo de esta sesión es trabajar en el ordenador con documentos XML y DTDs. La práctica no es entregable, pero el resto de prácticas se basarán en ella, por lo que es necesario llevarla a cabo con anterioridad a realizar las próximas prácticas.

Cada pareja debe definir un tipo de documento XML que permita representar un conjunto de películas que cumpla los requisitos que se enumeran a continuación:

- Como mínimo debe permitir representar los siguientes datos de una película, aunque se permite

representar cualquier otro tipo de datos adicional que el alumno considere conveniente:

- Título.
- Director.
- Año.
- Duración.
- Actores protagonistas (añadiendo, opcionalmente, el nombre del personaje o personajes cuyo papel desempeñen).
- Uno o más géneros a los que pertenezca.
- Nacionalidad o nacionalidades.
- Debe contar con un DTD que especifique los elementos permitidos, su tipo de contenido, atributos, etc. Dado el fin didáctico de la práctica, el DTD debe incorporar tipos de contenidos de elementos y de atributos variados.

Además de escribir el DTD, deben escribirse varios documentos XML con datos de películas. Deberá comprobarse que estos documentos estén bien formados y sean válidos.

Para comprobar si un documento dado está bien formado y es válido con respecto a su DTD se pueden utilizar los propios *parsers* de XML incorporados en la API de Java. El siguiente programa Java permite hacer esto: [Validate.java](#).