

a) Busca y analiza cuáles son las características propias del lenguaje XML e identifica la estructura de un documento con sus reglas sintácticas.

Características:

- Extensible: se pueden definir nuevas etiquetas. Permite la creación de etiquetas propias y permite asignar atributos a las etiquetas.
- Versátil: separa contenido, estructura y presentación.
- Estructurado: se pueden modelar datos a cualquier nivel de complejidad.
- Validable: cada documento se puede validar frente a un DTD/Schema.
- Abierto: independiente de empresas, sistemas operativos, lenguajes de programación o entornos de desarrollo.
- Sencillo: fácil de aprender y de usar.
- XML se almacena en formato texto (no binario) lo cual hace que los documentos sean directamente entendibles. Es decir, los documentos tienen una estructura entendible tanto por los ordenadores como por las personas.
- Orientado a los contenidos no presentación.
- No es sustituto de HTML.
- No existe un visor genérico de XML.
- Cada documento incluye metadatos sobre sí mismo, lo cual facilita la tarea de los motores de búsquedas en la web, ya que devolverán respuestas más adecuadas y precisas.
- Permite la exportabilidad a otros formatos de publicación de datos (HTML, PDF, texto enriquecido RTF, entre otros).
- XML es un estándar abierto no sujeto a ningún tipo de licencia
- XML permite la internacionalización, es decir puede trabajar con cualquier conjunto de caracteres, entre ellos el juego de caracteres UNICODE (utf-8).
- XML utiliza reglas de generación concretas y, por tanto, los documentos son fácilmente procesables.
- XML permite compartir información entre sistemas o fuentes de datos heterogéneas, por ejemplo, páginas web, distintas bases de datos, entre otras.

Identifica la estructura:

Un documento XML consta de un prólogo y un elemento raíz. El prólogo contiene la información sobre el resto del documento, como son la versión de XML y el código de caracteres utilizados. También en él se incluye la descripción de estructura del resto del documento, generalmente recogida en un DTD. Un documento XML tiene dos estructuras, una lógica y otra física. Físicamente, el documento está compuesto por unidades llamadas entidades. Una entidad puede hacer referencia a otra entidad, causando que ésta se incluya en el documento. Cada documento comienza con una entidad documento, también llamada raíz. Lógicamente el documento está compuesto de declaraciones, elementos, comentarios, referencias a caracteres e instrucciones de procesamiento, todos los cuales están indicados por una marca explícita.

Por otra parte, los elementos de un documento XML deben seguir una estructura de árbol, estrictamente jerárquica. Los elementos deben estar correctamente anidados. Los elementos no se pueden superponer entre ellos. Sólo puede haber un elemento raíz, en el que estén contenidos todos los demás.

¿Por qué es importante que los documentos XML estén bien formados y qué beneficios tiene para su posterior procesamiento?

Un documento XML está bien formado cuando cumple las reglas de creación de documentos, es decir, todos los procesadores de XML rechazarán cualquier documento que no esté bien formado. Cualquier restricción no cumplida será detectada y se tratará como un error fatal, el procesador informará a la aplicación y dejará de trabajar de una manera normal. Fundamentalmente, el objetivo de estas restricciones es asegurar que los documentos XML puedan ser interpretados por los procesadores XML sin ninguna ambigüedad, de manera que todos los elementos y atributos quedan perfectamente definidos. La validación XML, se encarga de comprobar que un documento en lenguaje XML está bien formado y se ajusta a una estructura definida.