

¿POR QUÉ ES NECESARIO VALIDAR UN DOCUMENTO XML?

Para tener un control más preciso sobre el contenido de los documentos es necesario un proceso de análisis exhaustivo.

La **VALIDACIÓN** es la parte más importante dentro de este análisis, ya que determina si un documento creado se ciñe a las restricciones descritas en el esquema utilizado para su construcción.

Controlar el diseño de documentos a través de esquemas aumenta su grado de fiabilidad, consistencia y precisión, facilitando su intercambio entre aplicaciones y usuarios, es decir, cuando creamos documentos XML válidos aumentamos su funcionalidad y utilidad.

La validación se encarga de verificar:

- La corrección de los datos: aunque validar contra un esquema no garantiza al 100% que los datos son correctos, nos permite detectar formatos nulos o valores fuera de rango y por tanto incorrectos.
- La integridad de los datos: al validar, se comprueba que toda la información obligatoria está presente en el documento.
- El entendimiento compartido de los datos: a través de la validación se comprueba que el emisor y receptor perciban el documento de la misma manera, que lo interpreten igual.

El éxito de la validación dependerá de factores como:

- Dónde se originan los documentos: si son fuentes confiables o no.
- Quién los crea: si son creados por una aplicación automáticamente o por un usuario de forma manual. (véase Herramientas)
- Quién los manipula: también es posible introducir errores involuntariamente durante la manipulación de los datos y documentos.
- La calidad de los datos: si los documentos se generan directamente de una base de datos de herencia, pueden no estar completos o correctos al 100%.
- El rendimiento del procesador o aparato que realice la validación: el procesado no es inmediato, necesita su tiempo. Si el rendimiento es crítico, se pueden aplicar diversas alternativas para reducir el coste computacional como limitar la validación a algunos aspectos, o crear un código específico para la aplicación particular que lo utiliza y se ejecute de forma más eficaz.

Documentos Válidos

Son aquellos documentos que además de estar bien formados, cumplen las especificaciones de la DTD (Document Type Definition), del Schema, o del elemento que lo valide, y siguen las pautas marcadas por sus modelos de contenido. No hay esquemas de documentos idénticos, por eso es difícil concretar las reglas que deben cumplir.

Para que un documento XML sea válido, durante el proceso de validación se comprueba:

- Qué elementos o atributos se permiten en un documento del tipo definido en el esquema.
- La estructura de los elementos y atributos (elementos anidados, atributos obligatorios u opcionales, etc.)
- El orden de los elementos.
- Los valores de los datos de atributos y elementos (según enumeraciones, rangos de valores delimitados, tipo de dato correcto (por ejemplo, formato correcto de una fecha, utilizar un entero para expresar un número), etc.
- La unicidad de valores dentro de un documento (por ejemplo, referencias de productos que no pueden repetirse).

Métodos de validación

Existen varios métodos para validar los documentos XML. Los métodos más usados son la **DTD** de XML versión 1.0, el **XML Schema** de W3C, **RELAX NG** de Oasis y **Schematron** de la Academia Sinica Computing Centre.

DTD

La DTD es el formato de esquema nativo (y el más antiguo) para validar documentos XML, heredado de SGML. Utiliza una sintaxis no-XML para definir la estructura o modelo de contenido de un documento XML válido:

Define todos los elementos.

- Define las relaciones entre los distintos elementos.
- Proporciona información adicional que puede ser incluida en el documento (atributos, entidades, notaciones).
- Aporta comentarios e instrucciones para su procesamiento y representación de los formatos de datos.

Es el método más sencillo usado para validar, y por esta razón presenta varias limitaciones, ya que no soporta nuevas ampliaciones de XML y no es capaz de describir ciertos aspectos formales de un documento a nivel expresivo.

Las DTD pueden ser internas o externas a un documento XML, o ambas cosas a la vez.

Uso de XML Schema

XML Schema es la evolución de la DTD descrita por el W3C, también llamado de forma más informal XSD (XML Schema Definition). Es un lenguaje de esquema más complejo y más potente, basado en la gramática para proporcionar una potencia expresiva mayor que la DTD. Utiliza sintaxis XML, cosa que le permite especificar de forma más detallada un extenso sistema de tipos de datos. A diferencia de las DTDs, soporta la extensión del documento sin problemas.

A la hora de validar, la utilización de XSD supone un gran consumo en recursos y tiempo debido a su gran especificación y complejidad en la sintaxis (esquemas más difíciles de leer y escribir).

Después de validar el documento con XML Schema, es posible expresar su estructura y contenido en términos del modelo de datos usado por el esquema de validación. Esta funcionalidad, conocida como Post-Schema-Validation Infoset (PSVI), se puede utilizar para transformar el documento en una jerarquía de objetos, a los cuales se puede acceder a través de un lenguaje de programación orientada a objetos (POO). El modelo de datos de XML Schema incluye:

- El vocabulario (nombres de elemento y atributo)
- El contenido modelo (relaciones y estructura)
- Tipos de datos.

Herramientas (EDITORES)

La creación manual de documentos XML e incluso su manipulación pueden introducir todo tipo de errores, tipográficos, sintácticos y de contenido. Existen editores de XML que facilitan la tarea de crear documentos válidos y bien formados, ya que pueden advertir de los errores básicos cometidos e incluso escribir automáticamente la sintaxis más sencilla necesaria.

Cuando necesitamos obtener un documento válido, el editor XML ha de ser capaz de:

- Leer la DTD del documento y presentarle una lista desplegable con los elementos disponibles enumerados en la DTD, evitando así la inclusión de algún elemento no definido en el esquema.
- Advertir el olvido de una etiqueta obligatoria e incluso no permitir este tipo de descuidos o errores, no dando por finalizado el documento si existen errores de este tipo.

XML COPY EDITOR

<https://xml-copy-editor.sourceforge.io/>