

UT.3. LENGUAJES DE MARCAS PARA EL ALMACENAMIENTO Y LA TRANSMISIÓN DE LA INFORMACIÓN. INTRODUCCIÓN AL XML.

Índice de contenido

1ORÍGENES DEL XML.....	1
2DEFINICIÓN DE XML.....	1
3UNA FAMILIA DE TECNOLOGÍAS.....	2
4SOFTWARE XML.....	4
5CARACTERÍSTICAS DE XML.....	5

1 ORÍGENES DEL XML

El XML proviene de un lenguaje que inventó IBM allá por los años 70. El lenguaje de IBM se llama GML (General Markup Language) y surgió por la necesidad que tenían en la empresa de almacenar grandes cantidades de información de temas diversos. Se necesitaba un lenguaje con el que poder clasificarlo todo y escribir cualquier documento para que se pueda luego procesar adecuadamente.

Este lenguaje gustó a la gente de ISO (International Organization for Standardization), una entidad de normalización para los procesos del mundo actual, de modo que allá por el 86 trabajaron para normalizar el lenguaje, creando el SGML, que no era más que el GML pero estándar (Standar en inglés).

SGML es un lenguaje muy trabajado, capaz de adaptarse a un gran abanico de problemas y a partir de él se han creado otros sistemas para almacenar información.

En el año 89, para el ámbito de la red Internet, un usuario Tim Berners-Lee (considerado el inventor de la World Wide Web) que había conocido el lenguaje de etiquetas (Markup) y los hiperenlaces creó un nuevo lenguaje llamado HTML, que fue utilizado para un nuevo servicio de Internet, la Web. Este lenguaje fue adoptado rápidamente por la comunidad y varias organizaciones comerciales crearon sus propios visores de HTML y se inició una encarnizada lucha para hacer el visor más avanzado, inventándose nuevas etiquetas sin orden ni concierto. Desde el 96 hasta hoy una entidad llamada W3C(World Wide Web Consortium) ha tratado de poner orden en HTML y establecer sus reglas y etiquetas para que sea un estándar. Sin embargo HTML no solucionó todos los problemas que planteaba la sociedad global de Internet.

Así que en el año 98, W3C empezó el desarrollo de XML (Extended Markup Language). El lenguaje pretendían solucionar las carencias del HTML en lo que se respecta al tratamiento de la información.

2 DEFINICIÓN DE XML

eXtensible Markup Language - Lenguaje de Marcado Extensible

XML es un metalenguaje, es decir que permite definir otros lenguajes de marcado adecuados a usos determinados (por ejemplo el lenguaje XHTML se construyó con XML).

XML es una tecnología muy sencilla que tiene a su alrededor otras tecnologías que la complementan y la hacen mucho más grande y con unas posibilidades enormes y fundamentales para la sociedad de la información. Por tanto deberíamos hablar de una familia de tecnologías XML en lugar de lenguaje XML (que es el término ampliamente usado).

La principal novedad de XML consiste en permitir compartir los datos con los que se trabaja a todos los niveles, por todas las aplicaciones y en todos los soportes. Esto es de vital importancia un mundo globalizado y que requiere la compatibilidad entre los sistemas.

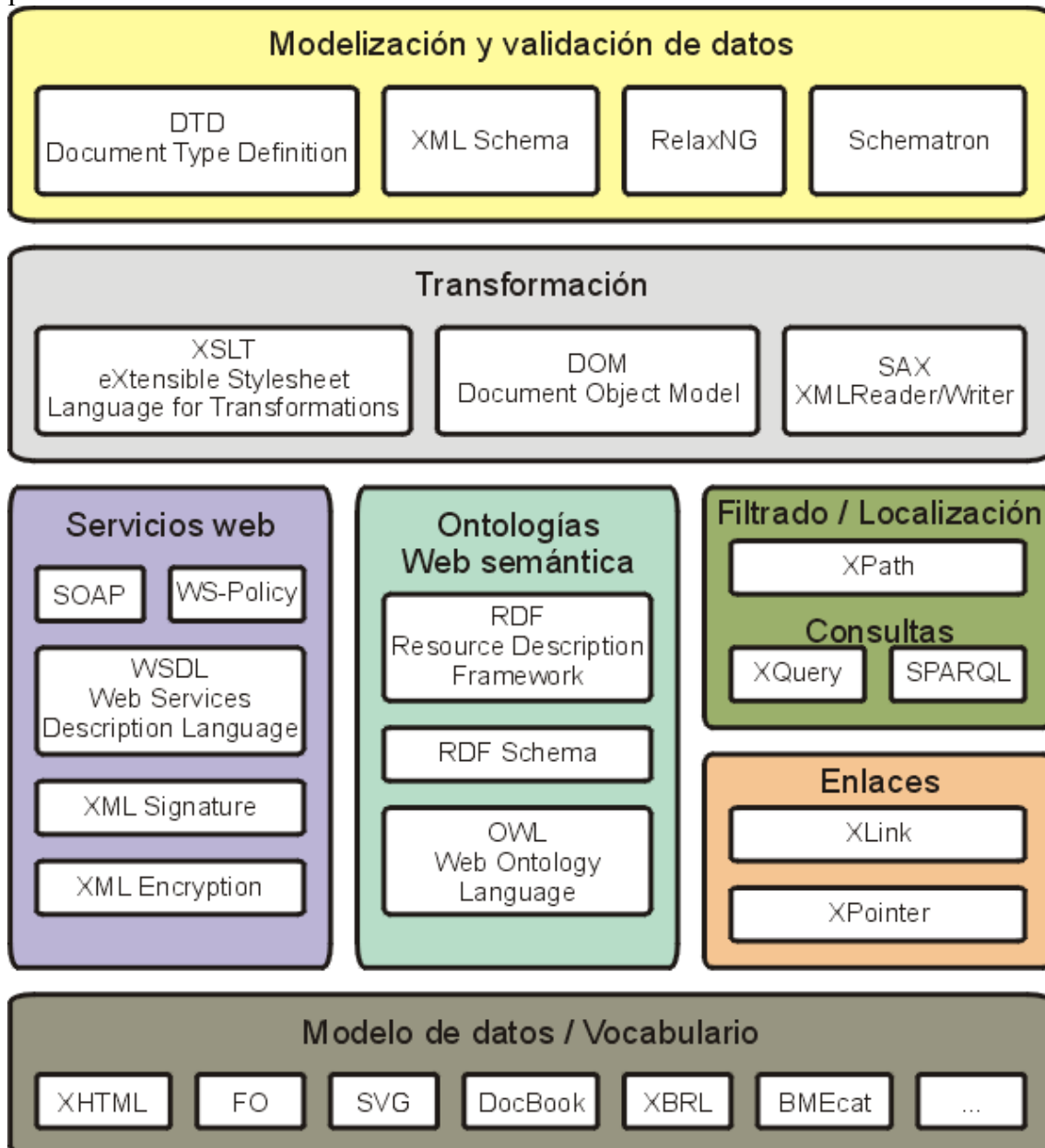
XML es la tecnología que permite compartir la información de una manera segura, fiable, fácil.

Con XML a los datos se les otorga significado, tienen sentido.

3 UNA FAMILIA DE TECNOLOGÍAS

Su objetivo consiste en crear una serie de tecnologías que trabajen juntas, originando un conjunto extensible de controles de formato de documentos que permita la reutilización de definiciones de documento para cualquier combinación de documentos y necesidades.

La siguiente imagen nos muestra la estructura de componentes/recomendaciones que presenta XML:



Algunas de las tecnologías más importantes son:

- **XML (versión 1.0):** Identifica los requisitos de un documento XML bien formado, así como el origen y los objetivos del XML. [<http://www.w3.org/XML>]
- **DTD: (Document Type Definition)** Una definición de tipo de documento contiene las reglas por las que es posible validar la información de un documento XML.
- **XML Schemas:** permiten incorporar más funcionalidad en los documentos XML de la que proporcionan actualmente las DTD.

- **XBase y XLink** (*XML Linking Language*): anteriormente conocido como **XLL**, define la forma en que los documentos XML deben enlazarse entre sí (hiperenlaces) [<http://www.w3.org/TR/xlink/>]
- **XPointer**: describe cómo debe apuntarse a un lugar específico dentro de un documento (anclas) [<http://www.w3.org/TR/xptr/>]
- **XPath**: proporciona un medio estándar de hacer referencia a partes dentro de documentos XML así como de manipulación y comparación de cadenas, números y valores booleanos. [<http://www.w3.org/TR/xpath/>]
- **XSL** (*eXtensible Stylesheet Language*): lenguaje la creación de hojas de estilo avanzado para XML. Se compone de: Un conjunto de propiedades y controles de formato para presentar la información de los documentos XML.
- **XSLT** (*XSL Transformations*) que define la sintaxis y la semántica que se utiliza para convertir documentos XML (permite reordenar, añadir y eliminar elementos y atributos)
- **XQL** (*XML Query Language*): proporciona un modelo de datos para su utilización con documentos XML, así como un lenguaje de consulta y un conjunto de operadores completos para el modelo de datos. Es un modo flexible de consulta de datos contenidos en documentos.
- **DOM y SAX** son un conjunto estándar de llamadas a funciones que permiten manipular documentos XML (y HTML) desde programas.
- **XML Namespaces** describe cómo se puede asociar una URL con cada etiqueta y atributo de un documento XML.
- **XEncryption, XSignature y XSecurity** tratan los aspectos de seguridad y autenticación en el intercambio de documentos XML.
- **XHTML** (*eXtensible Hypertext Markup Language*): es una reformulación de la especificación 4.0 de HTML existente en forma de módulos que siguen las reglas impuestas por la especificación XML.

..... Y MAS

4 SOFTWARE XML

Los documentos XML disponen de un conjunto de herramientas software que permiten la manipulación, creación, edición, presentación y análisis sintáctico de los documentos. Podemos dividir el software asociado a XML en tres grupos: editores, analizadores e interfaces de programación.

- Los **editores** nos facilitan la creación de los documentos XML. Se puede utilizar un editor tan sencillo como el bloc de notas, notepad++ o alguno más sofisticado, como Komodo o Amaya de W3C.
- Los **analizadores** nos proporcionan una herramienta capaz de entender la sintaxis de un documento XML y realizar comprobaciones sobre su estructura. Existen dos tipos básicos:
 - Los **no validadores** que sólo verifican que el documento XML está "bien formado". Expat, Lark, XP, TclXML, XML Testbed, RUWF

- Los **validadores** que realizan las tareas adicionales de verificación del contenido de un documento XML respecto a una DTD. LT XML, XML4Java, MS XML, MXL Checker
- Las **interfaces de programación de aplicaciones** (API, *Application Programming Interface*) proporcionan métodos para poder manipular documentos XML desde lenguajes de programación o de script. Dentro de este grupo cabe destacar el **DOM** (*Document Object Model*)

5 CARACTERÍSTICAS DE XML

XML es un subconjunto simplificado de SGML (Standard Generalized Markup Language).

XML es:

- Extensible: se pueden definir nuevas etiquetas.
- Estructurado: se puede modelar datos a cualquier nivel de complejidad, y su definición está en una DTD, Document Type Definition.
- Validable: cada documento se puede validar frente a una DTD, o en su defecto, se puede declarar bien formado.
- Independiente del medio: para publicar contenidos en múltiples formatos.
- Independiente del fabricante y de la plataforma: para poder utilizar cualquier herramienta estándar.
- XML es fácil de aprender y de usar.
- Los documentos XML son fácilmente procesables y compartibles en Internet.
- XML es tan popular por que cualquier lenguaje creado con él puede ser analizado sintácticamente por un procesador tan pequeño que puede ser incluido en un navegador Web.
- XML es texto, con sintaxis legible para los humanos lo que permite que los expertos puedan depurar los ficheros XML fácilmente con herramientas universales como el vi, edit o cualquier otro editor de textos.
- XML puede hacer por los datos lo que Java hace por los programas: independencia de plataforma y de fabricante. Por tanto permite una publicación independiente del medio
 - XML y XLS pueden reemplazar a los procesadores de texto y a los formatos de publicación.
 - Capacidad de publicar en un formato internacionalizado y capaz de interoperar con todos los productos y todas las plataformas.
- XML, aunque se puede utilizar con aplicaciones aisladas, está concebido para trabajar en la Web y posiblemente sea la innovación más importante desde la aparición de Java. Gracias a XML, aparece una Web llena de información en estado puro, enriquecida, estructurada, e inteligente, que supondrá una verdadera revolución, un salto de gigante hacia la consecución de una verdadera sociedad de la información.
- XML y Java tienen características complementarias, y cuando se utilizan juntos constituyen una plataforma muy potente para la compartición y el procesamiento de datos y documentos.

- XML define datos y documentos de una manera abierta y neutral, pero necesita aplicaciones que los procesen.
- Java ofrece un entorno de computación homogéneo con código que se puede transportar a cualquier máquina virtual Java.
- XML y Java hacen posible que se puedan aplicar los principios de "escribe una vez, ejecuta donde quieras" para procesar datos y documentos.

Resumiendo, XML provee:

- Independencia de los datos respecto de las aplicaciones.
- Información sobre la información, un contexto, extremadamente útil en su procesamiento informático.
- Elementos para describir la estructura de un documento.
- Herramientas para organizar la información a nuestro antojo.
- Formas de garantizar que los datos cumplen ciertas restricciones.
- Métodos para visualizar la información. Una forma de variar al gusto la visualización de la información y capacidad para visualizar la información en función del contenido, de la semántica del texto.
- Capacidad para pasar información a las aplicaciones.
- Enlaces para los datos relacionados.