

Lenguajes de Marcas y Sistemas de Gestión de Información

XML y otras tecnologías: conceptos básicos

Evolución tecnológica hasta llegar a XML

“El entorno tecnológico que estamos tratando es el del **intercambio electrónico de información** entre distintos ordenadores: intercambio de información a través de medios telemáticos.”

Problema de partida: “incompatibilidad entre diferentes plataformas que no permitía compartir información de forma directa” (manifiesto de IBM en la década de los 60)

“la red de ordenadores era un almacén con mucho contenido pero sin orden ni estructura”.

IBM creó **GML** (Generalized Markup Language- Lenguaje de Marcado Generalizado), para la edición y manipulación de documentos en la red.

Evolución tecnológica hasta llegar a XML: tres problemas a solucionar

1.

Los programas **no soportaban una representación genérica del documento**, es decir no existía el concepto de documento universal, por lo que si dos máquinas hablaban distinto lenguaje (pertenecen a distintas plataformas o con distinto software en ejecución) no se podían comunicar fácilmente y eran necesarios filtros de transformación que, casi siempre, se traducen en pérdida de información y, siempre, en ralentización de la comunicación.

Evolución tecnológica hasta llegar a XML: tres problemas a solucionar

2.

Cada aplicación tenía su **propio vocabulario**, independiente del resto de las aplicaciones, porque cada nuevo desarrollo planteaba la necesidad de comunicación y al no existir lenguaje alguno de referencia se creaba uno nuevo.

Esto introdujo el concepto de identificación de tipo de documento. Una primera solución para diferenciar el tipo de aplicación a la que correspondía un documento eran las extensiones. Así, por ejemplo, la extensión “.doc” correspondía a un procesador de textos, “.txt” a un editor de textos o “.dat” a un fichero de datos.

Evolución tecnológica hasta llegar a XML: tres problemas a solucionar

3.

Los ordenadores no tienen capacidad, por sí mismos, de **comprender el contenido de la información** que ellos mismos almacenan.

Por ejemplo: si se quiere buscar información sobre un tema, en realidad se debe buscar una palabra que represente dicho tema, la máquina al buscar remitirá como resultado todo documento que contenga esa palabra. Para desagrado del usuario, habrá muchos documentos que no tenían relación con el tema deseado. Para solucionar el problema se dieron propiedades al documento, de tal forma que al realizar la búsqueda se le añadiría las propiedades del documento requerido por el usuario.

Evolución tecnológica hasta llegar a XML

Como solución a estos problemas se introdujo el concepto de marcado, **separando dentro de un mismo documento el contenido de la estructura del mismo** (concepto primordial)

Nota: denominación de los lenguajes de marcado XXXML, donde **XXX** es el contenido concreto del lenguaje y **ML es Markup Language**; p.ej.: XML, HTML, GML, etc. (no confundir con otros xxxML como VRML – Virtual Reality Modelling Language).

Forma primordial de marcado

```
<estructura> contenido del documento </estructura>
```

donde observamos dos tipos o niveles de información: información que es el objeto del documento (lo que se desea almacenar) y la información estructural (identificar y estructurar lo almacenado); el límite entre los tipos está en los caracteres “<” y “>”.

Evolución tecnológica hasta llegar a XML

Una vez definido el marcado en los documentos, lo único que faltaba era conseguir que ese marcado fuese universal: la **estandarización del lenguaje**.

En la década de los 70 apareció **SGML** (Lenguaje de Marcado Generalizado Estándar), una versión de GML aceptado como estándar por ISO.

Dificultad para imponerse como estándar debido a su complejidad.

En 1989 el Laboratorio Europeo de Física de Partículas (CERN) creó un lenguaje de marcado con el objetivo de poder compartir documentos entre sus equipos de trabajo a través de **Internet**; era una aplicación sencilla y simple de SGML que se llamó ...

Evolución tecnológica hasta llegar a XML

... HTML

La idea inicial de HTML consistía en describir cómo eran los documentos que se deseaba intercambiar y que una herramienta que se llamó navegador, decidiera cómo se debía mostrar dicho documento.

Los documentos eran hiperdocumentos ya que contenían capacidades hipertextuales mediante las cuales se podía crear saltos o enlaces entre documentos (navegación).

HTML se popularizó y se inició una carrera por trasladar información a este formato y por dotar a esta tecnología de más y más posibilidades, como multimedia (imagen, sonido y otros) y por mayor interactividad de la información mostrada.

HTML se expandió de forma rápida pero desorganizada que finalmente se utilizó para fines no previstos en sus orígenes y que lo convirtieron en un lenguaje orientado a la presentación de la información al usuario alejándolo de la idea original de representación de información simple e independiente del navegador. La maraña de información creada así, la más grande y desorganizada que nunca ha existido ha recibido el nombre de ... **WWW** (Word Wide Web, Telaraña de ámbito mundial, Web o W3).

Evolución tecnológica hasta llegar a XML

Recientemente, se está utilizando HTML como **interfaz universal** para todo tipo de aplicaciones por lo sencillo que resulta y la obtención automática de millones de clientes capaces de manejar ese interfaz.

<Características y funcionamiento de Internet>

Evolución tecnológica hasta llegar a XML

HTML se ha quedado pequeño para ciertas aplicaciones, y sobre todo, se ha hecho necesario dotar a la Web de una estructura más semántica que permita organizar la telaraña. Por todo esto, el Consorcio de la Word Wide Web (W3C) intentó introducir SGML en la Web, con la intención de solucionar los problemas existentes con el HTML; el W3C buscó una alternativa que tuviera las ventajas de SGML y la sencillez de HTML.

En 1998, el W3C hizo público la normalización del estándar **XML** (Lenguaje de Marcado eXtensible. Alternativa muy potente (no tanto como SGML) pero muy sencillo (no tanto como HTML).

XML es un **metalenguaje** (lenguaje para definir otros lenguajes) al igual que SGML pero más sencillo y pensado para solucionar los problemas de la Web.

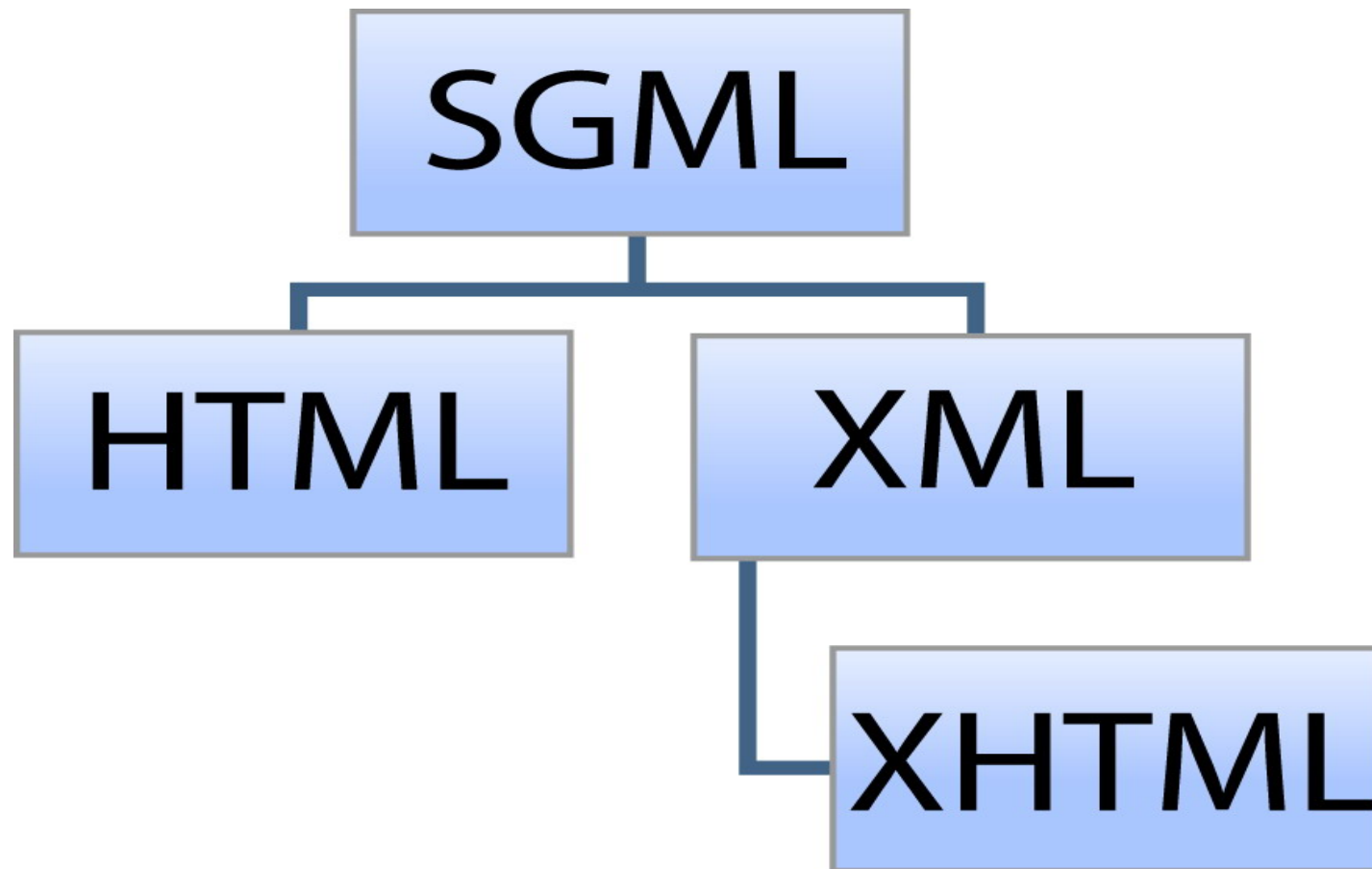


<http://www.w3.org/>



<http://www.xml.com>

Evolución tecnológica hasta llegar a XML



Objetivos del estándar XML

El estándar XML fue creado para que cumpliera los siguientes objetivos:

- Debía ser idéntico a la hora de servir, recibir y procesar la información que HTML, para aprovechar toda la tecnología implantada para el proceso de éste, es decir, debía ser compatible con HTML.
- Debía ser formal y conciso desde el punto de vista de los datos y la manera de guardarlos. Debía poseer las propiedades que poseía SGML, pero debía ser más fácil de utilizar que éste.
- Debía ser extensible, para que se pudiera utilizar en todos los campos del conocimiento realizando las extensiones precisas o usando los mecanismos necesarios.
- Debía ser fácil de leer y editar de forma directa por los seres humanos.
- Debía ser fácil de implantar, programar y aplicar a los distintos sistemas.

Uso actual de XML

Con XML se puede gestionar una gran diversidad de contenidos de información y representarla con diversos formatos (el preferido será el formato electrónico ya que está pensado para este medio y la automatización que permite). Mediante las herramientas específicas de trabajo con XML, se consigue trabajar con cualquier tipo de contenido.

La totalidad de los sistemas usan XML como representación de intercambio de documentos.

Documentos: Edición Orientada a la Presentación (POP) versus Middleware Orientado al Mensaje (MOM)

Edición Orientada a la Presentación:

Este tipo de edición crea documentos preocupándose principalmente de la apariencia que van a tener hacia el exterior, es decir cómo van a ser recibidos por el usuario final. La dificultad de este enfoque es muy elevada en la actualidad: los documentos se visualizan sobre multitud de plataformas con muy distintas características (p.ej.: teléfono móvil vs estación de trabajo de diseño gráfico).

HTML es un lenguaje orientado a la presentación y sufre los problemas de la complejidad comentados.

Esta forma de creación de contenidos no es la más adecuada para la Web pero es la única que nos ofrece resultados de gran calidad desde el punto de vista de la presentación.

Documentos: Edición Orientada a la Presentación (POP) versus Middleware Orientado al Mensaje (MOM)

Middleware Orientado al Mensaje:

En este caso se pretende crear documentos de intercambio entre aplicaciones. Con estos documentos se crea una capa intermedia de interacción entre sistemas que está alejada del usuario. En esta capa lo importante es la **representación semántica** de la información.

La clave de esta forma de crear documentos es que el usuario no los va a manejar de forma directa y, por tanto, no es necesario añadir información para facilitar la comprensión del usuario o para hacer el documento agradable, cómodo o eficiente para dicho usuario. Éste era el enfoque inicial de HTML pero no lo es en la actualidad; en cambio, XML ha sido creado expresamente con esta finalidad.

Relación XML-POP-MOM

En un documento XML se combinan ambas ideas, tanto POP como MOM.

Con XML sólo se crea la parte “inferior” del documento (su semántica) y no su apariencia externa (documentos MOM).

Para su visualización existen componentes adicionales externos, las hojas de estilo, que mediante instrucciones concretas añaden la información relativa a la representación externa, creando documentos POP.

Con las hojas de estilo se suprime la incompatibilidad entre los sistemas ya que se pueden crear varias hojas de estilo para un mismo documento y así tendremos una hoja de estilo para cada tipo de plataforma: un documento XML único y dos hojas de estilo, una para la estación de diseño y otra para el dispositivo inalámbrico.

Con XML se crea una etapa entre el cliente y el servidor que podríamos llamar “etapa transitoria”. En esta etapa se negocian todos los parámetros y propiedades para que el cliente pueda ver la información desde cualquier tipo de máquina, esto se puede realizar gracias al marcado de los datos, ya que éste proporciona la información necesaria sobre el contenido de documento, y la presentación final se realiza en HTML u otro lenguaje de representación a partir de los datos en XML y las hojas de estilo.

Relación XML-POP-MOM

Además de mejorar la compatibilidad de los documentos se aportan otras como:

Extracción de datos del documento. Esta operación permite extraer fragmentos semánticos del documento, y sólo se puede realizar sobre documentos MOM ya que en documentos POP no se guarda información sobre lo que significan sus datos, sólo sobre cómo deben presentarse. Podemos decir que con documentos MOM no se pierde información, debido fundamentalmente al marcado semántico.

Publicación del documento a disposición de múltiples usuarios y en distintos formatos. Esto sólo es posible si el documento original es MOM y se combina con componentes POP, las hojas de estilo. Utilización del documento, o parte del mismo, para realizar operaciones de búsqueda, ordenación. Esto sólo es posible sobre documentos MOM por el marcado semántico como ya se ha comentado.

Caso de empleo de XML

La organización **IMS** (Instructional Management System) emplea XML para la forma en que debe representarse la información pedagógica contenida en los sistemas de aprendizaje e-Learning.

[La Web semántica](#)

Bibliografía:

Gutierrez Martínez, J.M. et al.; Estándar XML y Tecnologías asociadas; DanyPress, 2003.