

LENGUAJES DE MARCAS Y SISTEMAS DE GESTIÓN DE INFORMACIÓN

Bloque XML: UD1, Introducción y un poco de Hª

XML: antes de..

2

- Es libre y abierto.
- No es tan nuevo: Viene de años 70
- Ahora en plena evolución o caída..
- Documentos:
 - ▣ antes texto=documento.
 - ▣ Ahora: documento es un fichero electrónico para cualquier media (texto, video, audio..)
- Destinatarios del documento:
 - ▣ Personas que los “consumen”
 - ▣ Máquinas que los “consumen”, para presentarlos o combinarlos con otros, etc.

Concepto de “marca”

3

- Es una forma de añadir información sobre los datos a los propios datos.
- Insertamos símbolos o caracteres en un documento para indicar su estructura lógica o física, cómo la información se debería presentar
- El objetivo principal es separar el tratamiento (por ejemplo apariencia) del documento de los datos.
- Ahora este concepto no es muy familiar: hemos estudiado previamente HTML, un lenguaje de marcas,

XML

4

- XML **NO** es un lenguaje de programación
 - ▣ Es un lenguaje de marcas
 - ▣ Es un “metalenguaje”
- Un metalenguaje es un lenguaje usado para describir formalmente otro lenguaje.
- XML permite a los desarrolladores crear lenguajes especializados, pero todos con algo en común: respetan la especificación XML.

Historia (I)

5

- Hace mucho tiempo.. En los 60s.. Y ahora..
 - ▣ Los documentos incluían códigos de control para definir presentación del texto.
 - ▣ Eran formatos **propietarios, cerrados**, dependientes del programa que procesaba.
- Finales de los 60 algunos pedían
 - ▣ Formato estándar y común para publicaciones (p.e. libros)
 - ▣ Separar el contenido, la información, de la forma en la que se va a presentar
- GML de IBM, año 1971: Generalized Markup Language
 - ▣ Extensible
 - ▣ Deberían existir mecanismos para comprobar si eran válidos
 - ▣ Deberían ser “legibles” a las personas.
- Otras tendencias
 - ▣ DTP aplicaciones: TEX, LaTeX, etc. Parecido, pero orientados a presentación.

Historia (II): Ejemplo GML

6

:h2. Definitions:

:ol.

li.1. noun, a gem variety of corundum in transparent or translucent crystals of a color other than red; especially, a transparent rich blue

:li.2. noun, a gem of such corundum

:li.3. noun, a deep purplish blue color

:li.4. adjective, made of or resembling a sapphire gem

:li.5. adjective, having the color of a blue sapphire

:eol.

SGML: *Standard Generalized Markup Language*

7

- Lo define ANSI (American National Standards Institute) a partir de GML 1980-84.
- ISO(International Organization for Standardization): 1986, en colaboración con ANSI
- Se centra en relación estructura y contenido u separa
 - ▣ Estructura de los datos
 - ▣ Contenidos
 - ▣ Estilo
- Formaliza el concepto de DTD, un documento separado pero relacionado con el documento y que especifica la estructura del mismo.
- Tanto XML como HTML derivan de SGML

Índice

8

- Introducción
- Un poco de historia
- GML, SGML, HTML, XML

¿y por qué no SGML?

9

- Es un estándar enorme, muy complejo
- Los lenguajes SGML son enormes y consumen demasiados recursos. Navegadores no lo soportan.
- Por ejemplo, añadir SGML a un procesador de texto dobla o triplica el precio del mismo..

HTML: antes de XML

10

- HyperText Markup Language (HTML)
 - ▣ desarrollado en el CERN (the European Organization for Nuclear Research).
- Propósito original: facilitar mecanismos independientes de plataforma para documentos hipertexto, con enlaces entre ellos, ya sea en el mismo servidor, red de área local o Internet.
- Internet ya era una tecnología madura, soportada por empresas, universidades, gobiernos, y HTML fue de gran ayuda para la difusión de contenidos de una forma sencilla y rápida.

HTML

11

- “Gratis”, muy simple, uso muy difundido, ampliamente soportado
 - ▣ => se hizo muy popular.
 - ▣ => Permitía a casi cualquiera tener una página web y presencia en la red.
- Estas ventajas lo hacen al mismo tiempo deficiente: no se puede extender, no hay estructura semántica, las empresas añaden etiquetas “propietarias”, etc.
- La mayoría de los diseñadores de páginas web solo se preocupaban de la apariencia del documento.

Vamos a arreglarlo..

12

- 1996: el W3C (World Wide Web Consortium) patrocinó a un grupo de expertos en SGML para desarrollar un lenguaje de marcas con la **potencia** y extensibilidad de **SGML**, pero con la **simplicidad** de **HTML**.
- Eliminaron todo aquello que no fuera básico, que no se usara, etc., => un lenguaje más pequeño, mas fácil y simple para implementar, orientado a la Web
- Así nace XML 1.0 alrededor de 1998.

Definición en la norma

13

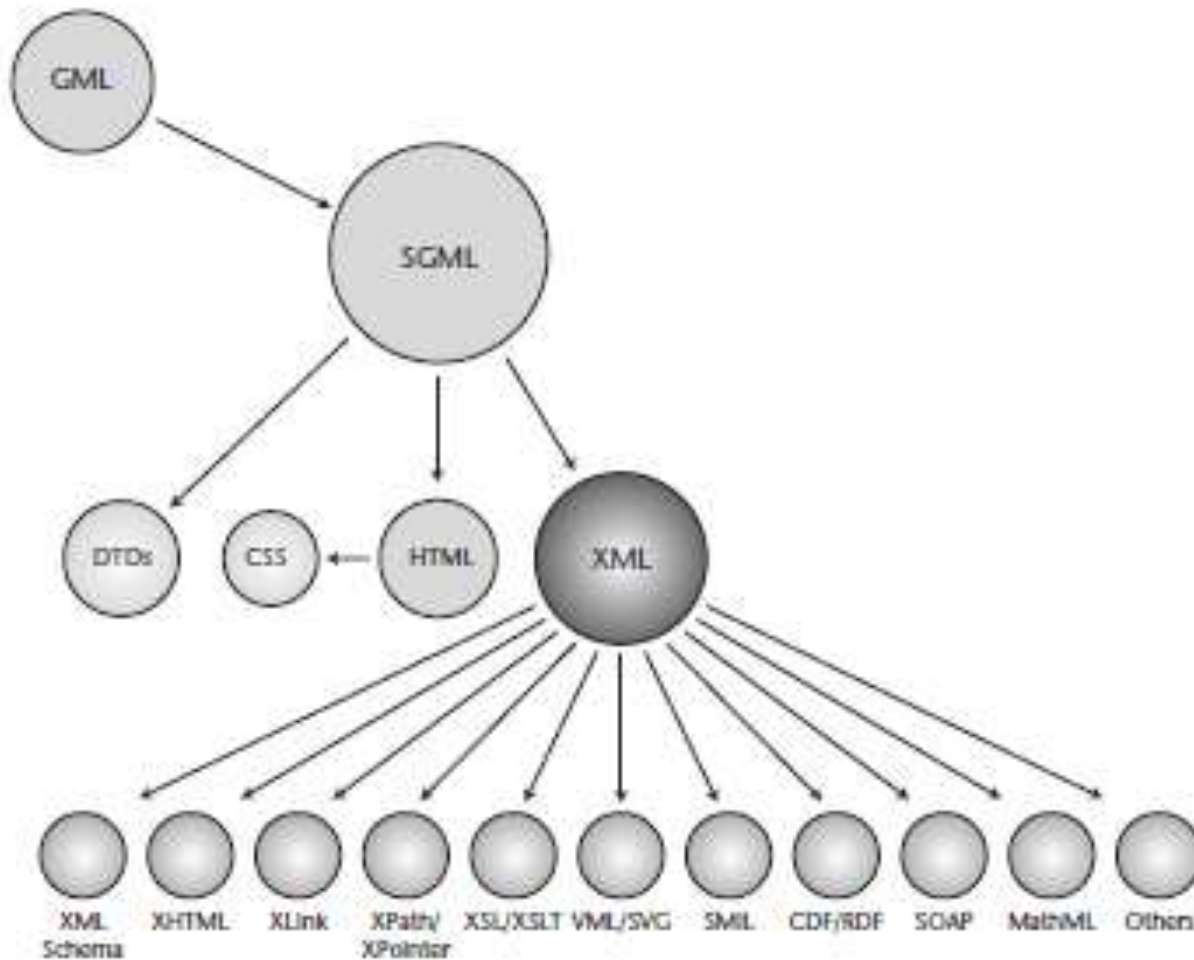
- The Extensible Markup Language (XML) is a subset of the Standard Generalized Markup Language. .
whose goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML.”

- Es un lenguaje de marcas y un metalenguaje
 - ▣ Actúa como estándar según el cual los desarrolladores pueden crear lenguajes específicos con sus propias etiquetas, estructuras, etc.
- No limita las marcas a crear, sólo impone que deben estar basadas en caracteres y ser legibles.
- Es independiente del SW y de la plataforma, así que un mismo documento XML puede ser usado en cualquier SO con cualquier herramienta de desarrollo o publicación.
- Es un formato público, un estándar abierto. No depende de empresa o país.
- Como es fácil de implementar, existen herramientas gratuitas.

- El **diseñador** se puede **centrar** en el uso de las etiquetas, en el **contenidos** y su **estructura**.
 - ▣ Formatear el contenido es un proceso posterior (e incluso a veces secundario)
- A los ficheros xml se les comprueba la sintaxis, y se puede comprobar la estructura (validación)
- Hay quien indica que debemos pensar en XML como en una versión menores de SGML, no como un HTLM++
 - ▣ XML se diseño pensando en la estructura y significado de los datos.
 - ▣ HTML, en cambio, con la forma en la que se presentan los datos.
- Las etiquetas XML describen datos y se usan como parte de transacciones de negocios, intercambios de información, etc.,
 - ▣ la sintaxis XML requiere precisión y seguir estrictamente las reglas gramaticales. Esto no ocurría con HTML hasta hace poco.

Gráficamente

16



Ejercicios

17

- XML en el PC: buscar ficheros con extensión xml
- Examinándolos
 - ▣ Abrirlos con editor de notas o navegador
- XML en la web: www.w3.org
 - ▣ Ver código fuente, posiblemente XHTML