

Extensible Markup Language

INTRODUCCIÓN



Antecedentes del XML

2

- Desde 1986 existe un estándar internacional que hace lo que XML pretende hacer, de hecho, hace más que XML en muchos aspectos. Su nombre es SGML (*Standard Generalized Markup Language*) o lo que es lo mismo ISO 8879.
- SGML se desarrolló para proporcionar un método de identificación de las partes y del contenido de un documento en base al tipo de información de su interior.
- SGML es un conjunto de reglas para la definición de lenguajes de marcado.



Antecedentes del XML

3

- SGML se crea para mantener almacenes de documentación estructurada en formato electrónico.
- SGML fue adoptado por la *Hacienda Pública* y el *DoD* de los EE.UU. Exigiendo que todos sus contratos lo incorporaran.
- Grandes corporaciones internacionales comienzan a utilizarlo como base para la estructuración de sus bases documentales.
- SGML fue utilizado por TBL como modelo para la creación del HTML.



El XML es...

4

- El Lenguaje de Marcas Extensible (*Extensible Markup Language, XML*).
 - ▣ Es un subconjunto propio de SGML.
 - ▣ Un metalenguaje de marcas.
 - ▣ Una sintaxis utilizada para crear lenguajes declarativos, un conjunto simple de reglas independientes para la representación de información textual estructurada.
- Una recomendación técnica del W3C (estándar en 1998).
- Multiplataforma, simple, fácil de aprender.
 - ▣ Es fácil construir herramientas para XML.
 - ▣ Optimizado para usarse en Internet.
- Libre (y gratuito).



El XML **no** es...

5

- Un lenguaje de marcas (*markup*).
 - ▣ **No**. Es un estándar que especifica una sintaxis para crear lenguajes de marcas.
- Solo para Web.
 - ▣ **No**. Puede ser usado para describir y comunicar cualquier información estructurada.
- Un superconjunto del HTML.
 - ▣ **No**. Aunque el HTML puede ser definido con sintaxis de XML (xHTML).
- Un invento de [x compañía].
 - ▣ **No**. XML es un estándar creado por el W3C y soportado por compañías e instituciones de todo el mundo.



El XML sirve para...

6

- Hacer publicación electrónica independiente del medio.
- Crear protocolos para el intercambio de datos entre miembros de una industria.
- Facilitar el procesamiento de datos usando software barato.
- Permite a las personas visualizar la información de la manera que quieran.
- Proporcionar metadatos que mejoran la calidad de la búsqueda de información.



XML versus HTML, SGML, PDF

7

- HTML: Lo bueno
 - ▣ El hipertexto funciona.
 - ▣ Es multiplataforma.
 - ▣ Tiene una curva de aprendizaje muy plana.
 - ▣ Barato (muchos editores, visores, verificadores, etc., gratuitos).
 - ▣ Base de información grande.
 - ▣ Los navegadores son baratos, sencillos de construir y de usar y poderosos.



XML versus HTML, SGML, PDF

8

- HTML: Lo Malo
 - ▣ Pobre herramienta de presentación.
 - Poco control de espaciado.
 - Problemas con el control de guiones, *kerning*, justificación y otras manipulaciones de texto.
 - EL uso de columnas es problemático.
 - ▣ Pobre herramienta de marcaje (*markup*).
 - No se pueden agregar etiquetas nuevas.
 - No es modular, poca oportunidad de reciclar.
 - Hay demasiado código inválido publicado actualmente.



XML versus HTML, SGML, PDF

9

- HTML: Lo peor
 - ▣ No puede ser extendido elegantemente.
 - Las etiquetas son fijas.
 - Las compañías y personas involucradas en hacer extensiones no saben de composición (*typesetting*) ni edición estructurada.
 - Es campo de batalla comercial (Mozilla vs IE).



XML versus HTML, SGML, PDF

10

- HTML: los problemas
 - ▣ La principal queja es:

“No puedo hacer [x cosa]”.

- ▣ Las razones:
 - Es solo un conjunto de etiquetas.
 - Pocos usan el DTD como guía.
 - Los navegadores siempre tienen que lidiar con código mal escrito.



XML versus HTML, SGML, PDF

11

- HTML: lo nuevo
 - ▣ Las hojas de estilo en cascada (*Cascading Style Sheets*, CSS).
 - ▣ Implementados en todos los navegadores en sus últimas versiones.
 - ▣ CSS1 emitida como recomendación del W3C en 1996.
 - ▣ CSS2 emitida como recomendación del W3C en 1998.
 - ▣ CSS2.1 recomendación en 2011
 - ▣ CSS3 (Color Module) emitida como recomendación del W3C en 2011.
 - ▣ Separa la estructura del formato.
 - ▣ Mayor control sobre la apariencia y posición.



XML versus HTML, SGML, PDF

12

- SGML: Lo bueno
 - ▣ Es multiplataforma.
 - ▣ Es un estándar ISO estable.
 - ▣ Hay disponibles muchas herramientas gratuitas para edición y conversión.
 - ▣ Es un conjunto de reglas, no un conjunto de etiquetas fijas.
 - ▣ Separa completamente la estructura del formato.



XML versus HTML, SGML, PDF

13

- SGML: Lo malo
 - ▣ Es complicado.
 - ▣ Es costoso.
 - El diseño de documentos es costoso.
 - La mano de obra es cara.
 - El entrenamiento es caro.
 - Aunque hay herramientas gratuitas, las que no lo son son muy costosas.



XML versus HTML, SGML, PDF

14

- SGML: Los problemas
 - ▣ La principal queja es:

¡El SGML es demasiado complicado y muy costoso!
 - ▣ Las razones:
 - El análisis de requerimientos es caro.
 - Los consultores son caros.
 - Las herramientas no son tan diversas ni tan flexibles.



XML versus HTML, SGML, PDF

15

- PDF: Lo bueno
 - ▣ Rápido y barato.
 - ▣ Preserva perfectamente la composición (*layout*) del documento.
 - ▣ Excelente para imprimir en cualquier dispositivo.
 - ▣ Multiplataforma.



XML versus HTML, SGML, PDF

16

- PDF: Lo malo
 - ▣ Archivos muy grandes.
 - ▣ Poca flexibilidad.
 - ▣ Pobres capacidades de búsqueda y navegación.
 - ▣ Pobre capacidad para reconvertir en otros formatos.
 - ▣ Pobre accesibilidad.



XML versus HTML, SGML, PDF

17

- PDF: Los problemas
 - ▣ La queja principal es:

¡Solo se puede imprimir!

- ▣ Razones:
 - Difícil de manipular.
 - La capacidades de los dispositivos diferentes.



XML versus HTML, SGML, PDF

18

- Se requiere algo nuevo
 - ▣ Barato, veloz y sencillo:
 - Para crear documentos.
 - Para procesar documentos.
 - Para presentar documentos.
 - ▣ Extensible:
 - Un conjunto de reglas, no un conjunto de etiquetas.
 - ▣ Compatible con el HTML:
 - Debe tener una manera sencilla de convertirse en HTML.
 - ▣ Compatible con el SGML:
 - Debe de conservar su potencia sin contener complejidades no necesarias.



Filosofía de XML

19

- La filosofía central de XML reside en la división del documento en sus tres componentes principales:
 - El **contenido**: la información del documento.
 - La **estructura**: el tipo y la organización de los elementos componentes del documento.
 - La **presentación**: la manera en que la información es presentada al lector.



Metas de diseño de XML

20

- XML será directamente utilizable en Internet.
- XML soportará una amplia variedad de aplicaciones.
- XML será compatible con SGML.
- Será fácil escribir programas que procesan documentos XML.
- El número de características opcionales en XML debe mantenerse en un mínimo absoluto, idealmente cero.
- Los documentos XML deberán ser legibles por humanos y resultar razonablemente claros.
- El diseño de XML deberá ser preparado rápidamente.
- El diseño de XML será formal y conciso.
- Los documentos XML serán fáciles de crear.
- El laconismo en las etiquetas XML es de importancia mínima.



Una familia de tecnologías

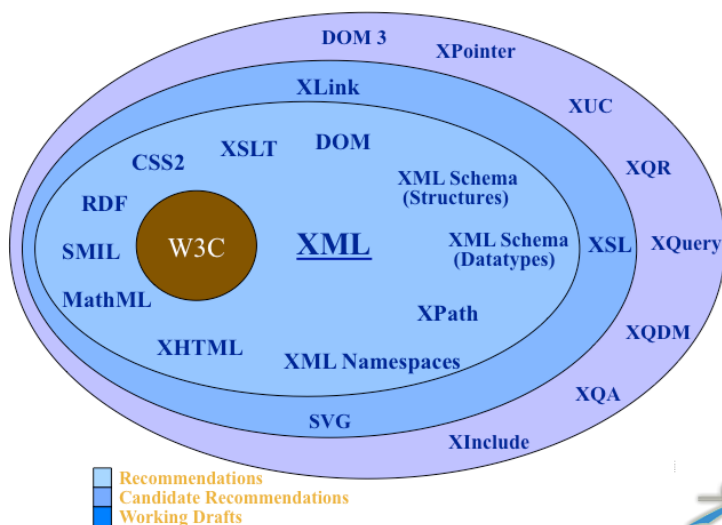
21

- 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.
- XML es una familia de tecnologías: XML, DTD, XLink, XPointer, XPath, CSS, XSL, XML Namespaces, XML Schemas, XQL, canonical XML, XHTML, XMLDS...



Una familia de tecnologías

22



Una familia de tecnologías

23

- Recommendations
 - XML (eXtensible Markup Language) v1.0 (Second Edition)
 - XML Schema Part1: Structures
 - XML Schema Part2: Datatypes
 - XHTML Basic (eXtensible HyperText Markup Language)
 - DOM Level 2 (Document Object Model)
 - XSLT (XSL Transformations)
 - Xpath (XML Path Language)
 - RDF (Resource Description Framework)
 - Namespaces en XML
 - SMIL (Synchronized Multimedia Integration Language)
 - CSS2.1 (Casacading Style Sheets)



Una familia de tecnologías

24

- Proposed and Candidate Recommendations
 - Xlink (XML Linking Language)
 - XSL (eXtensible Stylesheet Language)
 - SVG (Scalable Vector Graphics)
- Working Drafts
 - Xinclude (XML Inclusions)
 - Xpointer (XML Pointer Language)
 - DOM Level 3
 - Xquery (A Query Language for XML)
 - XML Query Requirements
 - XML Use Cases
 - XML Query Data Model
 - XML Query Algebra



Una familia de tecnologías

25

- **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:** Una definición de tipo de documento (*Document Type Definition*) contiene las reglas por las que es posible validar la información de un documento XML.



Una familia de tecnologías

26

- **XLink** (*XML Linking Language*, anteriormente conocido como XLL): define la forma en que los documentos XML deben enlazarse entre sí.

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

- **XPointer:** describe cómo debe apuntarse a un lugar específico dentro de un documento.

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

- **XPath:** proporciona un medio estándar de hacer referencia a direcciones y ubicaciones dentro de documentos XML así como de manipulación y comparación de cadenas, números y valores booleanos.

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



Una familia de tecnologías

27

- ▣ **CSS** (**C**ascading **S**tyle **S**heets): define un modelo de formato visual con controles avanzados para medios paginados e impresos.

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

- ▣ **XSL** (**eX**tensible **S**tylesheet **L**anguage): es un lenguaje diseñado específicamente para la creación de hojas de estilo con páginas XML. Se compone de:
 - Un conjunto de propiedades y controles de formato para presentar la información de los documentos XML.
 - **XSLT** (**XSL** Transformations): define la sintaxis y la semántica que se utiliza para convertir documentos XML.

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



Una familia de tecnologías

28

- ▣ **XML Namespaces**: Proporciona las reglas para asociar nombres de elementos y atributos de XML a conjuntos de vocabularios identificados previamente.

<http://www.w3.org/TR/1999/REC-xml-names-19990114/>

- ▣ **XML Schemas**: permiten incorporar más funcionalidad en los documentos XML de la que proporcionan actualmente las DTD.

<http://www.w3.org/TR/xmlschema-0/>

<http://www.w3.org/TR/xmlschema-1/>

<http://www.w3.org/TR/xmlschema-2/>



Una familia de tecnologías

29

- ▣ **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.

<http://www.w3.org/TR/xmlquery-req/>

- ▣ **Canonical XML**: el XML canónico crea un subconjunto del documento XML original que se puede comparar con otro documento que utilice una sintaxis igual o similar.

<http://www.w3.org/TR/xml-c14n/>



Una familia de tecnologías

30

- ▣ **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.

<http://www.w3.org/TR/XHTML1/>

<http://www.w3.org/TR/XHTML-basic/>

- ▣ **XMLDS** (*XML Digital Signatures*): describe las reglas de procesado y la sintaxis para la presentación de firmas digitales dentro de documentos XML.

<http://www.w3.org/TR/xmlsig-core/>



Áreas de Aplicación

31

- Banca en Línea (*Online Banking*)
- Canales (*Push Technology*)
- Automatización Web (*Web Automation*)
- Publicación de Base de Datos (*Database Publishing*)
- Distribución de Software (*Software Distribution*)

Corporaciones Usuarias

- Microsoft, Netscape, Sun Microsystems, Adobe, IBM, Oracle, Hewlett-Packard...



Ejemplos de XML

32

- Dentro del mundo de las denominadas tecnologías ‘push’
CDF: *Channel Definition Format*
- Dentro del mundo del comercio electrónico
OTP: *Open Trading Protocol*
OFX: *Open Financial eXchange*
- Dentro del mundo científico
MML: *Matemathical Markup Language*
CML: *Chemical Markup Language*



Ejemplos de XML

33

- Dentro de la distribución de software
OSD: *Open Software Distribution*
- Dentro de la automatización Web
WIDL: *Web Interfaz Definition Language*
- Dentro de la localización en Web
OTI: *Open Tag Initiative* (localización)
- Dentro del intercambio de Datos
FFI: *Lotus Notes Flat File Initiative*



Push Technology (CDF)

34

```
<?XML version="1.0"?>
<CHANNEL HREF="http://www.iechannelguide.com">
  <SELF HREF="http://www.iechannelguide.com/guide.cd" />
  <ABSTRACT>
    The Microsoft Channel Guide showcases the top content partners in the
    industry and lets you search for channels from around the world. Use it to
    find and add great content to your channel bar.
  </ABSTRACT>
  <SCHEDULE>
    <INTERVALTIME DAY="1"/>
  </SCHEDULE>
</CHANNEL>
```



Online Banking (OFX)

35

```
<?XML version="1.0"?>
<STATEMENTREQUEST>
  <BANKACCOUNT>
    <BANKID>123456</BANKID>
    <ACCOUNTID>9999</ACCOUNTID>
    <ACCOUNTTYPE>CHECKING</ACCOUNTTYPE>
  </BANKACCOUNT>
</STATEMENTREQUEST>
```



Software Distribution (OSD)

36

```
<?XML version="1.0"?>
<SOFTPKG HREF="http://www.microsoft.com/ie/ie40"
  NAME="{89820200-ECBD-11CF-8B85-00AA005B4383}"
  VERSION="4,72,2106,8"
  STYLE="ActiveSetup">
  <IMPLEMENTATION>
    <LANGUAGE VALUE="sp"/>
    <PROCESSOR VALUE="x86"/>
    <OS VALUE="winnt"/>
  </IMPLEMENTATION>
```



Scientific Publishing (CML)

37

```
<?XML version="1.0"?>
<MOLECULE>
  <ATOMS>
    <LOCATIONS>
      <X COORDS VALUE="1.0 1.1 3.2 4.6"/>
      <Y COORDS VALUE="2.3 2.4 5.5 7.3"/>
    </LOCATIONS>
  </ATOMS>
  <ELEMENTS>C H H H C C C</ELEMENTS>
</MOLECULE>
```



Aplicaciones XML

38

- Advertising: [adXML](#) place an ad onto an ad network or to a single vendor
- Literature: [Gutenberg](#) convert the world's great literature into XML
- Directories: [dirXML](#) Novell's Directory Services Markup Language ([DSML](#))
- Web Servers: [apacheXML](#) parsers, XSL, web publishing
- Travel: [openTravel](#) information for airlines, hotels, and car rental places
- News: [NewsML](#) creation, transfer and delivery of news
- Human Resources: [XML-HR](#) standardization of HR/electronic recruiting XML definitions
- International Dvt: [IDML](#) improve the mgt. and exchange of info. for sustainable development



Aplicaciones XML

39

- Voice: VoxML markup language for voice applications
- Wireless: WAP (Wireless Application Protocol) wireless devices on the World Wide Web
- Weather: OMF Weather Observation Markup Format (*simulation*)
- Geospatial: ANZMETA distributed national directory for land information
- Banking: MBA Mortgage Bankers Association of America → credit report, loan file, underwriting...
- Healthcare: HL7 DTDs for prescriptions, policies & procedures, clinical trials
- Math: MathML (Mathematical Markup Language)
- Surveys: DDI (Data Documentation Initiative) "codebooks" in the social and behavioral sciences

<http://www.oasis-open.org/cover/xml.html#applications>



Iniciativas en Comercio Electrónico

40

- CommerceNet
 - eCo Framework XML specs. to support interoperability among e-businesses
 - Commerce One Common Business Library (CBL): set of business components, docs. In DTD, XDR, SOX
 - BizTalk Microsoft spec. based on XML schemas
 - cXML (Commerce XML) -- tag-sets for e-procurement into BizTalk
- Electronic Data Interchange (EDI)
 - RosettaNet Common format for online ordering
 - FpML (Financial products Markup Language): sharing of financial data (interest rate & foreign exchange products)
- Open Buying on the Internet (OBI)
 - OBI high volume b2b purchasing transactions over the Internet (Office Depot, Lockheed, barnesandnoble, AX...



Iniciativas en Comercio Electrónico

41

□ E-commerce and XML

- **VISA Invoices** The Visa Extensible Markup Language (XML) Invoice Specification provides a comprehensive list of data elements contained in most invoices, including: **Buyer/Supplier, Shipping, Tax, Payment, Currency, Discount, and Line Item Detail.**

□ B2B Integration

- **code360** XML-Broker is middleware software that manages XML based transactions
- **Bluestone XML Suite** Enables to develop and deploy e-commerce, electronic data interchange, application integration and supply chain management applications. Bluestone XML Suite products include: XML-Server, Visual-XML, XML-Contact and XwingML.
- **webMethods** Provides companies with integrated direct links to buyers and suppliers

