

Tema 1. Introducción

¿Qué es?

XML, acrónimo de eXtensible Markup Language, se ha convertido en el formato estándar de facto para la representación de datos en Internet.

El eXtensible Markup Language es un estándar desarrollado por el grupo W3C o World Wide Web Consortium para textos estructurados. Este grupo se dedica a definir estándares en los que las empresas de software se basan posteriormente para crear sus aplicaciones.

Alrededor de XML existen muchas tecnologías relacionadas, que nos permitirán compartir información de una manera segura y fiable.

XML, no se aplica únicamente a Internet sino que proporciona el camino para el intercambio de información a las empresas. Por eso éstas prefieren el formato XML a otros porque es más fácil de interpretar y manipular, se podría decir que es casi autodescriptivo. De esta forma, si se recibe una cadena de texto normal sin ninguna identificación y la tenemos que interpretar, resulta difícil saber qué es exactamente; pero si ese texto va acompañado de unas etiquetas, sí que podemos identificarla (transforma los datos en información descriptiva). Es un Meta Markup Language (metalenguaje de marcas) y es extensible, ya que no es un formato prefijado como HTML.

XML refuerza el impacto de Internet en el mundo empresarial ya que se integran de esta forma más rápidamente. Es decir, si queremos modificar la presentación o contenido de páginas se requiere de menos tiempo y recursos humanos utilizando XML que otro formato.

¿Cómo surgió XML?

Los primeros orígenes de XML se encuentran en un lenguaje creado por IBM para almacenar grandes cantidades de información. Fue alrededor de los años setenta, cuando IBM inventó el lenguaje General Markup Language (GML). Este lenguaje les ayudaba a guardar toda la información que generaban, principalmente, a clasificarla y ordenarla, de tal manera que posteriormente pudiera procesarse adecuadamente.

Debido a su éxito sobre el año ochenta y seis, la Organización Internacional para la Estandarización o ISO (*International Organization for Standardization*), trabajó sobre este lenguaje para estandarizarlo creando SGML (*Standard Generalized Markup Language*), capaz de adaptarse a un gran abanico de problemas. A partir de él se han creado otros sistemas para almacenar información.

En el año ochenta y nueve, Timothy John Berners-Lee creó el lenguaje HTML HyperText Markup Language (*Lenguaje de Marcas de Hipertexto*), es decir el lenguaje de marcado predominante para la construcción de páginas Web. Este lenguaje tuvo un gran éxito, ya que es un sistema de presentación de documentos muy sencillo.

El auge de Internet y la necesidad de realizar páginas Web cada vez más vistosas, ha provocado que HTML no cumpla con las necesidades de intereses comerciales, ya

que es un lenguaje prodigioso en su momento pero que en la actualidad, por sí solo, no nos permite realizar todas las aplicaciones que nos gustaría.

Por este motivo, los miembros del W3C, pensando en solucionar las carencias de HTML sobre el año 1996, comenzaron la gestación del lenguaje XML (Extended Markup Language).

La primera versión salió a la luz el 10 de febrero de 1998, siendo la primera definición: "Sistema para definir validar y compartir formatos de documentos en la Web. A partir de este momento tuvo un gran crecimiento y difusión en la Red.

Es decir, XML surge después del lenguaje HTML y aprovecha las innegables ventajas de este lenguaje pero permite realizar muchas más cosas. Es decir, el objetivo de XML es crear un lenguaje muy general que sirva para muchas cosas.

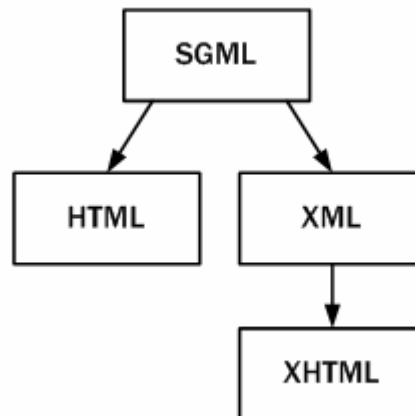
Mientras que HTML está diseñado para presentar información, principalmente, en cómo se muestra XML describe el contenido.

XML versus HTML

Como se ha expuesto en el apartado anterior, XML no es una nueva versión de HTML. Es cierto que ambos lenguajes, surgen del lenguaje SGML por lo que su sintaxis es similar, pero ambos tienen enfoques diferentes y, por tanto, cumplen diferentes funciones.

<i>HTML</i>	<i>XML</i>
Concebido para mostrar información. Su función radica en ayudarnos a darle formato a los diversos contenidos de una página.	Concebido para describir información. Su función principal es ayudarnos a organizar contenidos y eso hace que los documentos XML sean portables hacia diferentes tipos de aplicaciones.
Documentos desordenados.	Documentos ordenados.
No es case sensitive.	Es case sensitive.
Las etiquetas ya están todas definidas según el estándar definido por W3C.	Posibilidad de definir etiquetas propias.
Procesar la información en HTML es inviable, por estar mezclada con los estilos y las etiquetas que dan formato a la información.	En XML se puede procesar la información con mucha facilidad, porque todo está ordenado de una manera lógica.

Por lo expuesto hasta ahora, cabe pensar que HTML tenderá a desaparecer. Sin embargo, esto no es del todo cierto. Existe el lenguaje XHTML, que no es más que una adaptación de HTML al lenguaje XML. Técnicamente, HTML es descendiente directo del lenguaje SGML, mientras que XHTML lo es del XML (que a su vez, también es descendiente de SGML).



Las páginas y documentos creados con XHTML son muy similares a las creadas con HTML.

Características de XML

- XML es un estándar para escribir datos estructurados en un fichero de texto. Por datos estructurados entendemos documentos que van desde las hojas de cálculo, o las libretas de direcciones de Internet, hasta transacciones financieras o dibujos técnicos.
- XML parece HTML pero no lo es. En efecto, en XML se usan marcas y atributos, pero la diferencia estriba en que, mientras en HTML cada marca y atributo está establecido mediante un significado, incluyendo el aspecto que debe tener al verse en un navegador, en XML sólo se usan las marcas para delimitar fragmentos de datos, dejando la interpretación de éstos a la aplicación que los lee.
- XML está en formato texto, pero no para ser leído. Esto le da innumerables ventajas de portabilidad, depuración, independencia de plataforma, e incluso de edición, pero su sintaxis es más estricta que la de HTML: una marca olvidada o un valor de atributo sin comillas convierten el documento en inutilizable.
- XML no requiere licencias, es independiente de la plataforma, y tiene un amplio soporte. La selección de XML como soporte de aplicaciones, significa entrar en una comunidad muy amplia de herramientas y desarrolladores, y en cierto modo, se parece a la elección de SQL respecto a las bases de datos. Todavía hay que utilizar herramientas de desarrollo, pero la tranquilidad del uso del estándar y de su formato, hacen que las ventajas a la larga sean notables.
- Un fichero XML correctamente escrito requiere de dos condiciones:
 - Fichero XML bien formado es aquel que se ha escrito de acuerdo con el estándar.

- Fichero XML válido es aquel que -cumpliendo con la definición del estándar- está, además, lógicamente bien estructurado y define en su totalidad cada uno de sus contenidos sin ambigüedad alguna.

XML no es...

No hay que olvidar que XML es un lenguaje de marcas, ya que en ocasiones se puede pensar que XML es “la panacea” y la opción ideal para nuestra solución.

XML no es:

- Un lenguaje de programación, no existe un compilador que genere un programa ejecutable. XML se usa para enviar información de configuración u otro tipo para que el programa lo procese y tenga una reacción determinada.
- Un protocolo de transporte de red. XML por sí mismo no envía datos por la red. Es decir, los datos se envían por la red usando protocolos como http, ftp,.. y pueden estar codificados en XML.
- Una base de datos. XML no sustituye a Sistemas Gestores de Bases de Datos como, por ejemplo, Oracle o MySQL. Si es posible que en una base de datos se almacenen documentos XML pero la base de datos no es un documento XML

Usos de XML

Principalmente, XML se utiliza para el etiquetado y la estructuración de documentación, pero es tan potente y flexible que muchas personas lo utilizan para otros fines.

A continuación, se muestran algunos de los usos de XML:

➡ Para describir metacontenidos

Un metacontenido no es más que la información relativa al contenido del documento, como su título, autor, tamaño del archivo, fecha de creación, historial de cambios, palabras clave y demás información asociada.

Se puede utilizar un metacontenido, por ejemplo, para filtrar información o gestionar el documento.

XML está considerado el mejor vehículo para definir una sintaxis de metacontenidos dada su flexibilidad, legibilidad y capacidad de ampliación. Además, nos proporciona la posibilidad de definir estos metacontenidos fuera del documento.

Un ejemplo de formatos de metacontenidos definidos en XML es RDF:

- ✍ RDF (Resource Description Framework,): Marco de Descripción de Recursos es un marco de trabajo para metadatos en la World Wide Web, desarrollado por W3C.

➡ Para intercambiar contenidos de bases de datos

Muchas aplicaciones basadas en el modelo de tres capas extraen los datos a partir de sistemas remotos de bases de datos.

Si los datos se proporcionan en un documento XML que preserva la información original, como nombres de columnas y tipos de datos, será posible utilizarlos para otros fines, como por ejemplo, insertarlos en una hoja de cálculo para realizar operaciones.

➡ Para comunicar diferentes programas

Es sin duda alguna la aplicación más importante del XML, es decir, el intercambio de mensajes o información entre organizaciones o entre sistemas de aplicación dentro de una misma organización.

XML no está solo

XML consta de una familia de tecnologías. Por supuesto, existe una definición estándar de XML 1.0 de Febrero de 1998, pero su desarrollo se ha ido enriqueciendo paulatinamente a medida que se veían sus posibilidades: de esa forma, contamos con:

✍ CSS

CSS (Cascading Style Sheets), es decir, Hojas de Estilo en cascada es un lenguaje formal usado para definir la presentación de un documento estructurado escrito en HTML o XML (y por extensión en XHTML).

✍ XSL

XSL (Extensible Stylesheet Language), es decir, Lenguaje extensible de hojas de estilo es una familia de lenguajes basados en el estándar XML que permite describir cómo la información contenida en un documento XML cualquiera debe ser transformada o formateada para su presentación en un medio.

✍ XSLT

XSLT o Transformaciones XSL es un estándar de la organización W3C que presenta una forma de transformar documentos XML en otros e incluso a formatos que no son XML.

XSLT es muy usado en la edición Web, generando páginas HTML o XHTML. La unión de XML y XSLT permite separar contenido y presentación, aumentando así la productividad.

✍ XSL-FO

XSL-FO (eXtensible Stylesheet Language Formatting Objects) es un documento XML en el que se especifica cómo se van a formatear unos datos para presentarlos en pantalla, papel u otros medios. Figuran tanto los datos como el formato que se les va a aplicar.

✍ XLink

XLink (Lenguaje de vínculos XML) permite crear elementos de XML que describen relaciones cruzadas entre documentos, imágenes y archivos de Internet u otras redes.

XPointer

XPointer (Lenguaje de punteros XML) proporciona una forma de identificar de forma única fragmentos de un documento XML con el objetivo de realizar vínculos.

DOM

DOM (Document Object Model), es decir, Modelo en Objetos para la representación de Documentos es esencialmente un modelo computacional a través del cual los programas y scripts pueden acceder y modificar dinámicamente el contenido, estructura y estilo de los documentos HTML y XML.

JDOM

JDOM es una biblioteca de código fuente para manipulaciones de datos XML optimizados para Java

SAX

SAX (Simple API for XML), originalmente, una API únicamente para el lenguaje de programación Java, que después se convirtió en la API estándar de facto para usar XML en JAVA. Existen versiones de SAX no sólo para JAVA, si no también para otros lenguajes de programación.

XQuery

XQuery es un lenguaje de consulta diseñado para consultar colecciones de datos XML. Es semánticamente similar a SQL, pero incluye algunas capacidades de programación.

Xpath

XPath (XML Path Language) es un lenguaje que permite construir expresiones que recorren y procesan un documento XML. XPath permite buscar y seleccionar teniendo en cuenta la estructura jerárquica del XML.

Tema 2. Software de XML

Aunque más adelante existe un módulo sobre las herramientas de XML, en este tema se hará una breve descripción de las herramientas necesarias.

- Exploradores XML: se utilizan para ver e imprimir documentos XML.
- Editores XML: donde se escribirán los documentos XML. En la actualidad existe un rango muy amplio.
- Analizadores XML: en el caso de no querer “perder tiempo” con la sintaxis de XML, los programadores disponen de analizadores.

Cuando se trabaje con documentos XML realmente, como veremos más adelante, se trabajarán con *estructuras* de documentos. Será necesario abrir el fichero en el navegador de Internet, donde se pondrá en funcionamiento el *analizador o parser*. No será posible acceder a los contenidos hasta que el documento XML esté bien formado, es decir, cumpla las normas básicas de XML, las cuales se explicarán en el siguiente módulo.

Para escribir un documento XML, no es imprescindible ningún software y es suficiente con cualquier editor de texto que permita guardar los documentos en texto plano, como por ejemplo el Bloc de Notas de Windows.

En el módulo 4, Herramientas para trabajar con documentos XML, se detallarán los programas que tenemos a nuestra disposición para que nos faciliten la tarea de crear documentos XML, pero hasta entonces para los ejemplos del curso utilizaremos el editor básico que dispongamos en nuestro ordenador como, por ejemplo, el Bloc de Notas.

Tema 3. Ejemplos

A continuación, se muestra un ejemplo muy sencillo en XML, ya que aún no se ha explicado la sintaxis de XML pero que nos permitirá ir conociendo realmente cómo es XML.

Ejemplo típico: “Hola mundo”

Bueno, como no, en todos los manuales el primer ejemplo suele ser el “Hola Mundo”. Pues bien, mostraremos cómo sería en XML.

Abre el Bloc de Notas (u otro editor que te permita crear documentos de texto plano), escribe el código que viene a continuación y guarda el documento con la extensión .xml.

Código:

```
<?xml version="1.0" encoding="UTF-8"?>
<mensaje>
<parrafo>Hola Mundo</parrafo>
```

```
</mensaje>
```

Una vez que tengas el documento escrito, ábrelo desde el navegador de Internet. Si no has cometido ningún error obtendrás el siguiente resultado.

Resultado:

```
<?xml version="1.0" encoding="UTF-8" ?>
- <mensaje>
  <parrafo>Hola Mundo</parrafo>
</mensaje>
```

En ocasiones, al intentar visualizar un documento en el navegador se muestra un mensaje de error. Dependiendo del error cometido el contenido del mensaje puede variar, pero se mostrará algo similar a:

No se puede mostrar la página XML

No se puede ver la entrada XML con la hoja de estilo . Corrija el error y haga clic en el botón [Actualizar](#), o inténtelo de nuevo más tarde.

Sintaxis no válida para una declaración xml. Error al procesar el recurso file;

```
<?xml version="1.0" encoding="ISO-8859-1" >
-----^
```

Otro Ejemplo: Clasificación de los libros de una biblioteca.

A continuación, se mostrará un ejemplo algo más complejo que el anterior. Se trata de crear un documento XML con la información de los libros de una biblioteca determinada. Por ejemplo, representaremos los libros por su título, autor y categoría a la que pertenecen.

Igual que se hizo en el ejemplo anterior, abre el Bloc de Notas y escribe el siguiente código que correspondería a la solución XML que hemos propuesto.

Código:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
```



```
<biblioteca>
  <libro>
    <titulo> PORTA COELI: LA ORDEN DE SANTA CECLINA </titulo>
    <autor> SUSANA VALLEJO </autor>
    <categoria> Juvenil </categoria>
  </libro>

  <libro>
    <titulo> EL FUEGO </titulo>
    <autor> KATHERINE NEVILLE </autor>
    <categoria> Literatura negra y de intriga </categoria>
  </libro>

  <libro>
    <titulo> LA RESURRECCION MAYA </titulo>
    <autor> STEVE ALTEN </autor>
    <categoria> Literatura negra y de intriga </categoria>
  </libro>

  <libro>
    <titulo> CREPUSCULO </titulo>
    <autor> STEPHENIE MEYER</autor>
    <categoria> Juvenil </categoria>
  </libro>
</biblioteca>
```

A continuación, guarda el fichero con extensión .xml y abre el documento en el navegador de Internet.

Resultado:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <biblioteca>
-   <libro>
      <titulo>PORTA COELI: LA ORDEN DE SANTA CECLINA</titulo>
      <autor>SUSANA VALLEJO</autor>
      <categoria>Juvenil</categoria>
    </libro>
-   <libro>
      <titulo>EL FUEGO</titulo>
      <autor>KATHERINE NEVILLE</autor>
      <categoria>Literatura negra y de intriga</categoria>
    </libro>
-   <libro>
      <titulo>LA RESURRECCION MAYA</titulo>
      <autor>STEVE ALTEN</autor>
      <categoria>Literatura negra y de intriga</categoria>
    </libro>
-   <libro>
      <titulo>CREPUSCULO</titulo>
      <autor>STEPHENIE MEYER</autor>
      <categoria>Juvenil</categoria>
    </libro>
  </biblioteca>
```

Funcionamiento de XML

Como se ha mostrado en los ejemplos anteriores, se pueden editar documentos XML en un editor de texto plano cualquiera, sin necesidad de emplear una herramienta específica.

Los programas que hacen algo más que simplemente tratar al documento XML como cualquier otro archivo de texto, van a interpretar el contenido. Es decir, van a utilizar un analizador sintáctico XML para leer el documento, el cual es el responsable de dividir el documento en elementos (los cuales se explicarán en el siguiente módulo, pero brevemente te adelantamos que un elemento es cada etiqueta del documento XML) y otras piezas individuales.

Si en algún momento el analizador detecta un error, es decir, no se cumplen las reglas de una estructura bien formada de un documento XML, detiene el análisis. A veces, puede seguir leyendo el documento, para detectar e informar de otros errores.

Con estos dos ejemplos se ha mostrado cómo XML nos permite organizar y estructurar la información, a través de las etiquetas que nosotros mismos creemos. A partir de ahora se mostrará cómo utilizar XML para conseguirlo.

Tema 4. Recuerda Que...

XML, eXtensible Markup Lenguaje, es un estándar desarrollado por el grupo W3C o World Wide Web Consortium para textos estructurados

XML es un estándar para escribir datos estructurados en un fichero de texto. Por datos estructurados entendemos documentos que van desde las hojas de cálculo, o las libretas de direcciones de Internet, hasta transacciones financieras o dibujos técnicos.

XML parece HTML pero no lo es. En efecto, en XML se usan marcas y atributos, pero la diferencia estriba en que, mientras en HTML cada marca y atributo está establecido mediante un significado, incluyendo el aspecto que debe tener al verse en un navegador, en XML sólo se usan las marcas para delimitar fragmentos de datos, dejando la interpretación de éstos a la aplicación que los lee.

XML consta de una familia de tecnologías: CSS, XSL, XSLT , XSL-FO, XLink, XPointer, DOM, JDOM, SAX , XQuery , Xpath...