

CURSO 2022-2023
CICLO SUPERIOR DE DESARROLLO DE APLICACIONES WEB
IES AGUADULCE

Lenguajes de Marcas y Sistemas de Gestión de la Información

Francisco Javier Sueza Rodríguez

6 de octubre de 2022

Índice general

1	Lenguajes de Marcas y Sistemas de Gestión de la Información	4
1.1	Definición y Clasificación de los Lenguajes de Marcas	4
1.2	Evolución de los Lenguajes de Marcas	5
1.2.1	El origen: GML y SGML	5
1.2.2	La Popularización: HTML	6
1.2.3	La Madurez: XML	7
1.2.4	Comparación XML y SGML	7
1.2.5	Comparación XML y HTML	8
1.3	Etiquetas, Elementos y Atributos	8
1.4	Herramientas de Edición	9
1.5	XML	10
1.5.1	Estructura y Sintaxis	10
1.5.2	El Prólogo	11
1.5.3	El Ejemplar	12
1.5.3.1	Elementos	13
1.5.3.2	Atributos	14
	Glosario	16
	Bibliografía	17

Índice de figuras

1.2.1 Documento SGML simple	6
1.2.2 Documento HTML simple	6
1.2.3 Documento XML simple	7
1.3.1 Partes de un elemento HTML	9
1.5.1 Ejemplo del <i>ejemplar</i> en un documento XML	12
1.5.2 Código XML sin indentación (incorrecto)	12
1.5.3 Código XML con indentación (correcto)	13
1.5.4 Uso de atributos en documentos XML	14

Tema 1

Lenguajes de Marcas y Sistemas de Gestión de la Información

En esta unidad vamos a estudiar los aspectos básicos de los lenguajes de marcas y los sistemas de gestión de la información. Por un lado, veremos la evolución de los **lenguajes de marcas**, desde GML hasta HTML, así como sus elementos y atributos, haciendo especial énfasis en XML. A continuación, veremos en que consisten los **sistemas de gestión de la información**, en concreto los **ERP**, sus características, configuración básica, personalización,..etc.

1.1 Definición y Clasificación de los Lenguajes de Marcas

Los «lenguajes de marcas» sirven para **codificar un documentos**. Estos incorporan **etiquetas** o marcas con **información adicional** sobre como se estructura el texto o como se presenta. El lenguaje de marcas será el que defina que etiquetas se permiten, donde deben colocarse y que significado tienen.

Todo lenguaje de marcas esta definido en un documento denominado **DTD**, donde se establecen las marcas, los elementos utilizados por dicho lenguaje y sus correspondientes etiquetas y atributos, así como su sintaxis.

Los lenguajes de marcas se pueden clasificar, principalmente, en tres grupos:

- **Orientados a la presentación:** son los utilizados generalmente por los procesadores de texto y definen como debe presentarse el documento, es decir, el formato que tiene.
- **De procedimientos:** orientados también a la presentación, pero en este caso, dentro de un **marco procedural** que permite la definición de macros, es decir, el programa que representa el documento debe interpretar el código en el mismo orden que aparece. Algunos ejemplos son **TeX**, **LaTeX** y **Postscript**
- **Descriptivos o semánticos:** estos lenguajes no describen la presentación del documento, sino que **describen la información**, que es lo que se esta representando sin especificar como debe presentarse.

Algunos ejemplos de lenguajes de marcado agrupados por su ámbito de uso son los siguientes:

- **Documentación Electrónica:**
 - **RTF (Rich Text Format):** fue desarrollado por Microsoft en 1987 y permite el intercambio de documentos entre los diferentes procesador de texto.

- **TeX**: creado por **Donald Knuth**, este lenguaje está especialmente enfocado en la creación de textos científicos. Es considerado la mejor forma de componer fórmulas matemáticas complejas. [1]
 - **Wikitexto**: permite la creación de páginas wiki en servidores preparados para soportar este lenguaje.
 - **DocBook**: permite generar documentos separando la estructura lógica del documento de su formato, permitiendo que estos documentos puedan publicarse en diferentes formatos sin tener que modificar el documento original.
- **Tecnologías de Internet:**
 - **HTML, XHTML** (Hypertext Markup Language, eXtensible Hypertext Markup Language): estos lenguajes están orientados a la creación de páginas web.
 - **RSS** (Really Simple Syndication): permite la difusión de contenido web mediante la sindicación de contenidos.
 - Otros lenguajes especializados:
 - **MathML** (Mathematica Markup Language): especializado en expresar los formalismos matemáticos de forma que puedan ser entendidos por diferentes aplicaciones.
 - **VoiceXML** (Voice eXtended Markup Language): permite el intercambio de información entre usuarios y una aplicación con capacidad de reconocer el habla.
 - **MusicXML**: permite el intercambio de partituras entre diferentes editores de partituras.

1.2 Evolución de los Lenguajes de Marcas

A finales de los **años 60** surgen unos lenguajes informáticos, diferentes de los lenguajes de programación, orientados a la gestión de la información. Con el desarrollo de los editores y procesadores de texto surgen los primeros lenguajes informáticos orientados a la descripción y estructuración de la información: **los lenguajes de marcas**. Paralelamente también surgen otros lenguajes orientados a la representación, almacenamiento y consulta de grandes cantidades de datos: lenguajes y sistemas de bases de datos.

Los lenguajes de marcas surgieron inicialmente como lenguajes formados por un conjunto de códigos que los procesadores de textos insertaban en los documentos para dirigir el proceso de presentación (impresión) mediante una impresora. Al igual que los lenguajes de programación, estos estaban **ligados** a las características de los **procesadores de texto** y las **impresoras** en los que se usaban y no permitían a los programadores abstraerse de dichas características.

Posteriormente se añadió como medio de presentación a la pantalla y se automatizó el proceso, teniendo ya solo que pulsar una combinación de teclas para lograr los resultados deseados en vez de hacerlo a mano. Este marcado estaba orientado exclusivamente a la presentación de la información, aunque posteriormente se le dieron nuevos usos surgiendo con ello el **formato generalizado**.

1.2.1 El origen: GML y SGML

Uno de los problemas que ha tenido la informática ha sido la **falta de estandarización** en los formatos de información usados por los diferentes programas.

En los años 60, **IBM** encargó a **Charles F. Goldfarb** la construcción de un sistema de edición, almacenamiento y búsqueda de documentos legales. Después de analizar el funcionamiento de la empresa se

llego a la conclusión de que necesitaban un formato estándar a todos los departamentos para gestionar la documentación.

Así fue como se creó **GML**, un formato que permitía describir los documentos de tal forma que el resultado fuese independiente de la plataforma o la aplicación utilizada. Este formato evolucionó hasta que en 1986 se creó el estándar **ISO 8879** donde se especificaba el formato **SGML**, un lenguaje muy complejo y que requería de unas herramientas de software caras, por lo que su uso quedó relegado a grandes aplicaciones industriales.

```
<email>
  <remitente>
    <nombre>Peter</nombre>
    <apellido>Pan</apellido>
  </remitente>
  <destinatario>
    <direccion>campanilla@paisdenuncajamas.com</direccion>
  </destinatario>
  <asunto>Paseo</asunto>
  <mensaje>¿Te apetece dar una vuelta?</mensaje>
</email>
```

Figura 1.2.1: Documento SGML simple

1.2.2 La Popularización: HTML

En 1990, **Tim Berners-Lee** creó el World Wide Web y conociendo SGML, se encontró con la necesidad de compatibilizar, enlazar y organizar gran cantidad de documentos procedentes de diversos sistemas. Como solución, a partir de la sintaxis de SGML, creó un lenguaje de descripción de documentos llamado **HTML**, combinando dos estándares ya existentes:

- **ASCII**: código basado en el alfabeto latino, tal como se usa en inglés moderno [2]. Cualquier procesador de textos simple puede reconocer y almacenar este formato, permitiendo la transferencia de datos entre dos ordenadores.
- **SGML**: lenguaje que permite dar estructura al texto aplicando diferentes formatos.

HTML es una **versión simplificada de SGML**, ya que solo utiliza las instrucciones absolutamente necesarias. Gracias a su simplicidad, tuvo un éxito rotundo en la World Wide Web, convirtiéndose rápidamente en el **estándar general** para la **creación de páginas web**. Actualmente, HTML es el tipo de documento más utilizado en el mundo.

```
<!DOCTYPE HTML>
<html>
  <head>
    <meta charset="utf-8" />
    <title>Ejemplo1</title>
  </head>
  <body>
    <p>Párrafo de ejemplo</p>
  </body>
</html>
```

Figura 1.2.2: Documento HTML simple

1.2.3 La Madurez: XML

Uno de los problemas que surgió con HTML es que la cantidad de documentos escritos en este lenguaje creció exponencialmente, muchos de los cuales no se ceñían a ningún estándar generando bastante caos. Como respuesta es ese problema, el **W3C** estableció en 1998 el estándar internacional **XML**, un lenguaje de marcas puramente estructural, que **no incluye información sobre el diseño**, y permite la creación de etiquetas adaptadas a las necesidades, convirtiéndose con rapidez en el estándar para intercambio de datos en la web.

XML es un **metalenguaje** con las siguientes características:

- Permitir definir etiquetas propias.
- Permitir asignar atributos a las etiquetas.
- Utilizar un esquema para definir de forma exacta las etiquetas y sus atributos.
- La estructura y el diseño son independientes.

En realidad XML es un **conjunto de estándares** relacionados entre sí y que comprende los siguientes:

- **XLS** (eXtensible Style Language): permite definir hojas de estilo para XML e incluye capacidad de transformación de documentos.
- **XML Linking Language**: determina aspectos sobre los enlaces entre documentos XML e incluye **Xpath**, **Xlink** y **Xpointer**.
- **XML Namespaces**: proveen de un contexto donde se aplican las marcas del documento XML y que se diferencian de otras con el mismo nombre válidas en otros contextos.
- **XML Schemas**: permiten definir restricciones que se aplicarán a un documento XML. Actualmente las más utilizadas son **DTD**.

```
<?xml version="1.0" encoding="UTF-8x°x"?">
<!DOCTYPE biblioteca>

<biblioteca>
  <ejemplar tipo="libro" isbn="978-2-7460-4958-1" edicion="1">
    <titulo>XML practico</titulo>
    <editorial>Ediciones Eni</editorial>
    <autor>Sebastien Lecomte</autor>
    <autor>Thierry Boulanger</autor>
    <autor funcion="traductor">Ángel Belinchon Calleja</autor>
    <prestamos>
      <lector inicio="13/05/2014" devolucion="15/05/2014">Pedro López</lector>
      <lector inicio="13/07/2015" devolucion="15/07/2015">Ali Méndez</lector>
    </prestamos>
  </ejemplar>
</biblioteca>
```

Figura 1.2.3: Documento XML simple

1.2.4 Comparación XML y SGML

Aunque XML está basado en SGML, estos tienen muchas diferencias. A continuación se muestra una tabla con las principales diferencias de estos dos lenguajes de marcas.

XML	SGML
Su uso es sencillo	Su uso es muy complejo
Trabaja con documentos bien formados	Solo Trabaja con documentos válidos
Desarrollo de aplicaciones a bajo coste	Aplicaciones para procesarlo costosas
Muy utilizado en informática y otras áreas	Se utiliza en sectores muy específicos
Compatibilidad e integración con HTML	No hay compatibilidad con HTML
Formateo y estilos fáciles de aplicar	Formateo y estilo relativamente complejos

Como vemos en esta tabla, SGML es un lenguaje mas complejo y costoso que XML, además de imponer mas restricciones, haciéndolo un lenguaje menos flexible que XML y mas orientado a sectores concretos. Para obtener más información sobre XML, podemos consultar el [Estándar XML](#) publicado por la W3C.

1.2.5 Comparación XML y HTML

Aunque tanto XML como HTML se crearon a partir de SGML y su sintaxis es similar, son lenguajes diferentes con propósitos diferentes, como podemos ver en la siguiente tabla.

XML	HTML
Es un perfil de SGML	Es una aplicación de SGML
Permite definir conjuntos de etiquetas	Aplica un conjunto limitado de etiquetas
Modelo de hiperenlaces complejo	Modelo de hiperenlaces simple
Navegador como plataforma de desarrollo	Navegador como visor de páginas
Compatibilidad e integración con HTML	No hay compatibilidad con HTML
Fin de las etiquetas propietarias	Problema de la no compatibilidad en navegadores

1.3 Etiquetas, Elementos y Atributos

Los lenguajes de marcas usan una serie de etiquetas intercaladas en un documento sin formato, las cuales serán posteriormente por el interprete del lenguaje.

Existen tres términos ampliamente utilizados en los lenguajes de marcas:

- **Etiquetas:** una etiqueta, también llamada **tag**, es un pequeño bloque de código que se escribe encerrado entre los símbolos **menor que** (<) y **mayor que** (>). Normalmente se utilizan **dos etiquetas**, una de **inicio** y otra de **fin**, para indicar que el efecto que queríamos conseguir ha finalizado, con la única diferencia que a la etiqueta de fin se le añade el carácter / antes del nombre.
- **Elemento:** representan estructuras mediante las que se organiza el contenido del documento, o acciones que se desencadenan cuando el navegador lo interpreta. Está **compuesto** de la **etiqueta de apertura**, la **etiqueta de cierre** y el **contenido entre ambas**.

- **Atributo:** es un par **nombre-valor**, que se encuentra al inicio de un elemento e indican diferentes propiedades asociadas a ese elemento

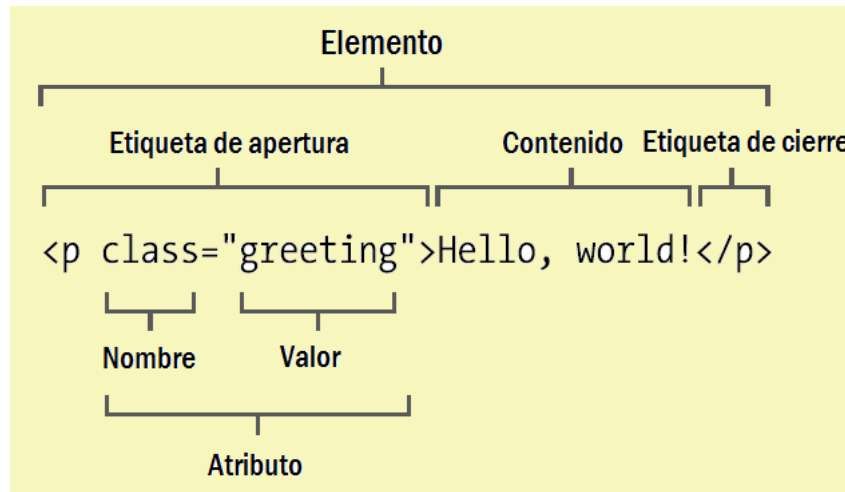


Figura 1.3.1: Partes de un elemento HTML

1.4 Herramientas de Edición

Para trabajar con XML es necesario, por un lado, editar los documentos, y por otro procesarlos. Por ello, necesitaremos dos tipos de herramientas para trabajar con él, estas son:

■ Editores XML

Una característica de los lenguajes de marcas es que se basan en la utilización de ficheros de **texto plano**, por lo que basta con usar cualquier editor de texto para construir un documento XML.

Aunque podemos usar cualquier editor, cuando elaboramos documentos XML complejos es conveniente usar algún software de edición XML. Estos nos ayudan a crear estructuras y etiquetas de los elementos usados en XML, resaltan las etiquetas para diferenciarlas más cómodamente y además incluyen ayudas para la creación de otros elementos como DTD, hojas de estilo CSS o XLS,.. El W3C ha desarrollado un editor HTML, XHTML, CSS y XML gratuito llamado Amaya, pero existen otros muchos gratuitos como: Notepad++, VSCode, Sublime Text, Netbeans,..etc

■ Procesadores XML

Para interpretar un documento XML puede usarse cualquier navegador. Los procesadores XML permiten visualizar los documentos XML y acceder a su contenido y estructura. Un **procesador** es un conjunto de módulos de software entre los que se encuentra un **parser**, que comprueba que el documento cumple las normas establecidas para que pueda abrirse. Estas normas pueden corresponderse a las necesarias para trabajar con documentos de tipo válido o solo exigir que el documento este bien formado. A los primeros se le conocen como **procesadores validadores** y a los segundos como **procesadores no validadores**.

Para publicar documentos XML en internet se usan **procesadores XSLT**, que permiten generar archivos HTML a partir de XML.

Puesto que XML se usa para el intercambio de archivos entre aplicaciones, hay que recurrir a motores independientes como «XML para Java» de IBM, JAXP de Sun, etc

1.5 XML

XML, que significa *eXtensible Markup Language*, es un lenguaje que permite definir lenguajes de marcas desarrollado por el W3C y utilizado para almacenar datos de forma legible. Proviene del lenguaje SGML y permite definir la gramática de lenguajes específicos para estructurar documentos. [3]

Su importancia radica en que permite **compartir datos** entre diferentes equipos y aplicaciones de forma **segura, fiable y sencilla**. El hecho de que diferentes equipos y aplicaciones puedan leer y generar archivos en este formato lo convierte en una herramienta muy útil para el envío de información a través de la Web. Aunque a veces suele confundirse con HTML, podemos decir que HTML está diseñado para mostrar datos en nuestras pantallas mientras que XML está diseñado para almacenar y compartir esos datos.

El XML ahorra tiempos de desarrollo y proporciona ventajas, dotando a webs y aplicaciones de un método muy potente para almacenar y compartir información. Por ello, se ha convertido en un formato universal usado por todo tipo de sistemas operativos y dispositivos.

1.5.1 Estructura y Sintaxis

Un documento XML es un documento de texto, con la **extensión .xml**, compuesto de un **conjunto de etiquetas, estructuradas en árbol**, que describen la organización del documento y que es interpretado por un navegador Web.

La **características básicas** de XML son las siguientes:

- Es directamente **compatible** con protocolos usados en la Web como **HTTP** y **URL**.
- Todo documento que verifique las reglas de XML está **conforme con SGML**.
- **No se requieren conocimientos de programación** para realizar tareas sencillas en XML.
- Los documentos XML son **fáciles de crear**.
- **La difusión** de documentos XML **está asegurada**, ya que cualquier procesador de XML puede leer documentos XML.
- El marcado de XML es **legible para los humanos**.
- El diseño de XML es **formal y conciso**.
- XML es **extensible, adaptable y aplicable** a una gran variedad de situaciones.
- XML es **orientado a objetos**
- Todo documento XML se **compone** de **datos de marcado** y **datos carácter** entremezclados.

El **proceso de creación** de un documento pasa por varias etapas en las que el éxito de cada una de ellas se basa en la calidad de la anterior. Estas etapas son:

1. Especificación de requisitos.
2. Diseño de etiquetas.
3. Marcado de los documentos.

El **marcado** son etiquetas que se añaden para estructurar los documentos y permitir a los ordenadores interpretar los textos. Los **datos carácter** son los que componen la verdadera información del documento.

Los documentos XML también pueden **contener comentarios**, que son interpretados por el intérprete XML y que comienzan con la cadena `<!--` y finalizan con `-->`. Pueden estar en cualquier posición del documento salvo en:

- El prólogo
- Dentro de una etiqueta

Los XML están **formados** por dos partes, **el prólogo** y **el ejemplar**, los cuales veremos a continuación.

1.5.2 El Prólogo

El prólogo es la primera parte que nos encontramos en cualquier documento XML y siempre debe preceder al ejemplar del documento. Éste se divide en dos partes, la **declaración XML** y la **declaración de la codificación** empleada, los cuales explicamos a continuación.

1. **La declaración XML:** es la primera línea del documento, de no ser así, se genera un error que impide que el documento sea procesado. En caso de que sea opcional, se permite el procesamiento de documentos HTML y SGML como si fueran documentos XML, si es obligatoria, estos deberán incluir la declaración de la versión XML. Esta declaración permite indicar de forma explícita que el documento es de tipo XML.

El prólogo puede tener **tres funciones**:

- a) **Declaración de la versión** utilizada de XML:

```
<?xml version= "1.0"?>
```

- b) **Declaración de la codificación** empleada:

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

En este caso se usa el código código iso-8859-1 (Latin-1) que permite el uso de tildes o caracteres como la «ñ». Otro de los códigos a tener en cuenta es **UTF-8 (unicode)**. Esta codificación soporta más caracteres que iso-8859-1 y permite que estos se visualicen correctamente en más sistemas. Es la **codificación estándar** recomendada para todos los documentos y la usaremos siempre, salvo que por algún motivo no pueda ser empleada.

Para consultar más información sobre codificación de caracteres en sistemas informáticos y cuales son los más empleados podemos visitar [página de Wikipedia](#) sobre codificación de caracteres.

- c) **Declaración de autonomía** del documento:

Indica si el documento necesita de otro para su interpretación. Para declararlo, hay que definir el prólogo completo:

```
<?xml version= "1.0" encoding="iso-8859-1" standalone="yes" ?>
```

La opción *standalone* indica al procesador XML si un documento es independiente (*standalone=yes*), o si depende de un documento externo, como declaraciones de marcas externas o una DTD externa (*standalone=no*). Por defecto el documento se considera independiente.

2. La declaración del tipo de documento:

Esta declaración define el tipo de documento que estamos creando para que sea procesado correctamente. Toda declaración de tipo de documento comienza con la cadena:

```
<!DOCTYPE Nombre_tipo ...>
```

1.5.3 El Ejemplar

Es la parte más importante de un documento XML ya que contiene los **datos reales** del documento. Es el **elemento raíz** de un documento XML y este se nombrará igual que la declaración del tipo de documento (*!DOCTYPE*). Suele estar compuesto de **elementos anidados**.

En el siguiente ejemplo, el ejemplar es el elemento `<libro>`, que a su vez está compuesto de los elementos `<titulo>`, `<autoria>`, `<editorial>`, `<isbn>`, `<edicion>` y `<paginas>`.

```
<?xml version="1.0" encoding="utf-8" standalone="yes"?>
<!DOCTYPE libro>

<libro>
  <titulo>XML práctico</titulo>
  <autoria>
    <autor>Sebastien Lecomte</autor>
    <autor>Thierry Boulanger</autor>
  </autoria>
  <editorial>Ediciones Eni</editorial>
  <isbn>978-2-7460-4958-1</isbn>
  <edicion>1</edicion>
  <paginas>347</paginas>
</libro>
```

Figura 1.5.1: Ejemplo del *ejemplar* en un documento XML

Es recomendable **establecer un criterio** y mantenerlo durante todo el documento, por ejemplo, que las etiquetas vayan escritas **siempre en minúsculas**.

Por otro lado, es conveniente anidar los elementos para una visualización óptima del documento, esto se hará **indentando** o **tabulando** el código. A continuación se muestran dos figuras, una sin indentación (errónea) y otra con indentación (correcta).

```
<?xml version="1.0" encoding="utf-8" standalone="yes"?>
<!DOCTYPE libro>

<libro>
<titulo>XML práctico</titulo>
<autoria>
<autor>Sebastien Lecomte</autor>
<autor>Thierry Boulanger</autor>
</autoria>
<editorial>Ediciones Eni</editorial>
<isbn>978-2-7460-4958-1</isbn>
</libro>
```

Figura 1.5.2: Código XML sin indentación (incorrecto)

```

<?xml version="1.0" encoding="utf-8" standalone="yes"?>
<!DOCTYPE libro>

<libro>
  <titulo>XML práctico</titulo>
  <autoria>
    <autor>Sebastien Lecomte</autor>
    <autor>Thierry Boulanger</autor>
  </autoria>
  <editorial>Ediciones Eni</editorial>
  <isbn>978-2-7460-4958-1</isbn>
</libro>

```

Figura 1.5.3: Código XML con indentación (correcto)

Como podemos observar, en la segunda figura se reconoce más fácilmente la estructura del documento y se facilita la lectura de este.

Por último, es recomendable **anidar grupos de datos relacionados**, como se ha hecho en el ejemplo anterior con los elementos `<autor>` dentro del elemento `<autoria>`, ya que el documento que **mas limpio** y **ordenado**.

1.5.3.1 Elementos

Todos los datos de un documento XML deben **pertenecer** a un **elemento del mismo**.

Los elementos son los diferentes **bloques de información** que permiten definir la **estructura** del documento XML. Estos están delimitados por una **etiqueta de apertura** y una **etiqueta de cierre**. Pueden estar **compuestos** por **otros elementos**.

Los **nombres** de las etiquetas deben ser **autodescriptivos**, es decir, que ilustren su contenido. Por ejemplo, si estamos con datos relativo a un libro, una etiqueta no debería ser `<caracteristicas>`, ya que es demasiado ambiguo, deberíamos utilizar nombres como `<titulo>`, `<isbn>`, etc...

La **formación de elementos** debe cumplir ciertas **normas** para que queden bien definidos y que el documento XML al que pertenecen pueda ser procesado sin generar ningún error fatal. Estas normas son las siguientes:

- En todo documento XML debe existir **un documento raíz** y **solo uno**.
- **Todos** los elementos tienen una **etiqueta de inicio** y **otra de cierre**. En caso de que en el documento existan **elementos vacíos**, se pueden sustituir las etiquetas de apertura y cierre por una de elemento vacío. Esta se construye como una etiqueta de inicio pero añadiendo `/`, es decir, `<elemento/>` en vez de `<elemento>`. Esta considerada una etiqueta de apertura y cierre.
- Al anidar hay que tener en cuenta que **no puede cerrarse** ningún **elemento** que **contenga otro elemento** que aún **no se haya cerrado**.
- Los **nombres** de las **etiquetas de apertura y cierre** deben ser **idénticos**, respetando las mayúsculas y minúsculas. Pueden ser cualquier cadena alfanumérica que no contenga espacios ni tildes, y que no comience por dos puntos (:), ni por la cadena **xml**.
- El contenido de los elementos **no puede contener** la cadena `</>`, por compatibilidad con SGML. Además, no se pueden utilizar los caracteres **mayor que** (`>`), **menor que** (`<`), **ampersand** (`&`), **dobles comillas** (`"`) y **apostrofe** (`'`).

En caso de tener que utilizar alguno de estos caracteres, se deben sustituir por las siguientes cadenas:

Caracteres	Cadena	Caracteres	Cadena
>	>	"	"
<	<	'	'
&	&		

- Para **utilizar caracteres especiales**, como £, ©, ®,... hay que usar las expresiones **&#D;** o **&#H;**, donde D y H se corresponden con el número decimal o hexadecimal, respectivamente, correspondiente al carácter que se quiere representar en código **UNICODE**.

En los siguientes enlaces puedes consultar tanto los códigos ASCII y su equivalente en HTML, como el código correspondiente a cada carácter en UNICODE.

- ASCII - <https://www.asciitable.com/>
- UTF-8 - <https://www.charset.org/utf-8>

1.5.3.2 Atributos

Las etiquetas pueden tener **atributos**, que **permiten definir propiedades** a los elementos de un documento. Los atributos, a diferencia de los elementos, no puede organizarse en ninguna jerarquía o estructura de árbol, no pueden contener ningún otro elemento, no pueden contener valores múltiples y no reflejan ninguna estructura lógica. En definitiva, los atributos **no se podrán extender fácilmente** en futuros cambios.

Un elemento puede tener varios atributos, pero ninguno de estos puede estar vacío. Los atributos se codifican de la siguiente forma:

```
<etiqueta atributo="valor_atributo"></etiqueta>
```

No debe usarse un atributo para almacenar información susceptible de ser dividida, sino para proporcionar información adicional sobre el elemento. A continuación vamos a ver un ejemplo de la utilización de atributos en un documento XML.

```
<?xml version="1.0" encoding="utf-8" standalone="yes"?>
<centro_educativo>
  <alumno sexo="Varón" fecha_nacimiento="05/06/1990">
    <nombre>Pablo</nombre>
    <apellido>Pérez</apellido>
    <telefono tipo="Móvil">666666666</telefono>
    <direccion tipo="Calle">Policar, 34</direccion>
  </alumno>
</centro_educativo>
```

Figura 1.5.4: Uso de atributos en documentos XML

Por norma general, intentaremos **evitar el uso** de atributos o procurar **no abusar de ellos**. Normalmente **los usaremos para metadatos** o información que no sea relevante para los datos. Se puede ver como una manera de incorporar características o propiedades a los elementos, como en el siguientes

ejemplo: `<perro raza="Pastor">Lolo</perro>`. También se suelen para especificar unidades de medida y similares: `altura unidad_altura="cm">178</altura>`. Hay que destacar que toda la información que se almacena en atributos se podría almacenar igualmente en elementos.

Lo que si es recomendable es que **una vez elegido un estilo**, mantenerlo dentro de todo el documento XML, teniendo en cuenta que los **nombres de los atributos** tienen que cumplir las **mismas normas** que los **elementos**, y no pueden contener el carácter menor que "<".

Glosario

ASCII American Standard Code for Information Interchange. 6, 11

DTD Document Type Definition. 4, 11

GML Generalized Markup Language. 6, 11

HTML Hypertext Markup Language. 6, 11

metalenguaje Lenguaje que permite la definición de otros lenguajes. 7, 11

SGML Standard Generalized Markup Language. 6, 11

texto plano Es aquel texto formado solo por datos sin formato, es decir, solo por caracteres.. 9, 11

unicode Es un código que permite el tratamiento informático de textos en cualquier lenguaje y disciplina técnica, ya que incluye todos los caracteres conocidos para cualquier lengua. Es compatible con ASCII. 11

UTF-8 8-bit Unicode Transformation Format. 11

W3C Word Wide Web Consortium. 7, 11

XML eXtensible Markup Language. 7, 11

Bibliografía

- [1] Wikipedia - Tex. <https://es.wikipedia.org/wiki/TeX>.
- [2] Wikipedia - ASCII. <https://es.wikipedia.org/wiki/ASCII>.
- [3] Wikipedia - XML. <https://es.wikipedia.org/wiki/XML>.