

Raúl Rubio Vázquez

# Características de los Lenguajes de Marcas

**UNIDAD 1**

- 1. Objetivos generales de la unidad*
- 2. Competencias y contribución en la unidad*
- 3. Contenidos conceptuales y procedimentales*
- 4. Evaluación*
- 5. Contenidos teóricos*

# 1. Objetivos generales de la unidad

1. Conocer las características básicas de los lenguajes de marcas.

3. Estudiar las características propias de la estructura de los documentos XML.

5. Conocer las ventajas que aportan los espacios de nombres.

2. Clasificar los lenguajes de marcas según sus ámbitos de aplicación.

4. Comprender las reglas sintácticas de la estructura de los documentos XML.

## 2. Competencias y contribución de la unidad

Lenguajes de Marcas

Unidad 1 - Características de los Lenguajes de Marcas

*h) Desarrollar interfaces gráficos de usuario interactivos y con la usabilidad adecuada, empleando componentes visuales estándar o implementando componentes visuales específicos.*

*En un primer acercamiento al módulo se muestran conceptos básicos de los lenguajes de marcado. Comenzando por el estudio de la formación correcta de un documento XML.*

# 3.Contenidos

CONTENIDOS CONCEPTUALES		CONTENIDOS PROCEDIMENTALES
Introducción	<ul style="list-style-type: none"><li>• Concepto de lenguaje de marcas.</li><li>• Ventajas de los lenguajes de marcado.</li><li>• SGML</li><li>• Organizaciones y estándares</li></ul>	<ul style="list-style-type: none"><li>• Comprender el significado de marca.</li><li>• Diferenciar los lenguajes de marcas de los lenguajes de programación.</li><li>• Vislumbrar las ventajas y conocer el origen de este tipo de lenguajes.</li></ul>
Características Y Clasificación	<ul style="list-style-type: none"><li>• Propiedades de los lenguajes de marcas.</li><li>• Clasificación de los lenguajes de marcas.</li></ul>	<ul style="list-style-type: none"><li>• Dominar las características propias de los lenguajes de marcas.</li><li>• Clasificar los lenguajes de marcas atendiendo al tipo de marcas que utilizan y a su funcionalidad.</li></ul>
Principales lenguajes de marcas	<ul style="list-style-type: none"><li>• HTML</li><li>• XML</li></ul>	<ul style="list-style-type: none"><li>• Diferenciar entre los lenguajes derivados de SGML según su utilidad.</li></ul>

# 4.Evaluación

*Se han identificado las características generales de los lenguajes de marcas*

*Se han reconocido las ventajas que proporcionan en el tratamiento de la información.*

*Se han clasificado los lenguajes de marcas e identificado los más relevantes.*

*Se han diferenciado sus ámbitos de aplicación.*

*Se han reconocido la necesidad y los ámbitos específicos de aplicación de un lenguaje de marcas de propósito general.*

*Se han analizado las características propias del lenguaje XML*

*Se han identificado la estructura de un documento XML y sus reglas sintácticas.*

*Se han contrastado la necesidad de crear documentos XML bien formados y la influencia en su procesamiento.*

*Se han identificado las ventajas que aportan los espacios de nombres.*

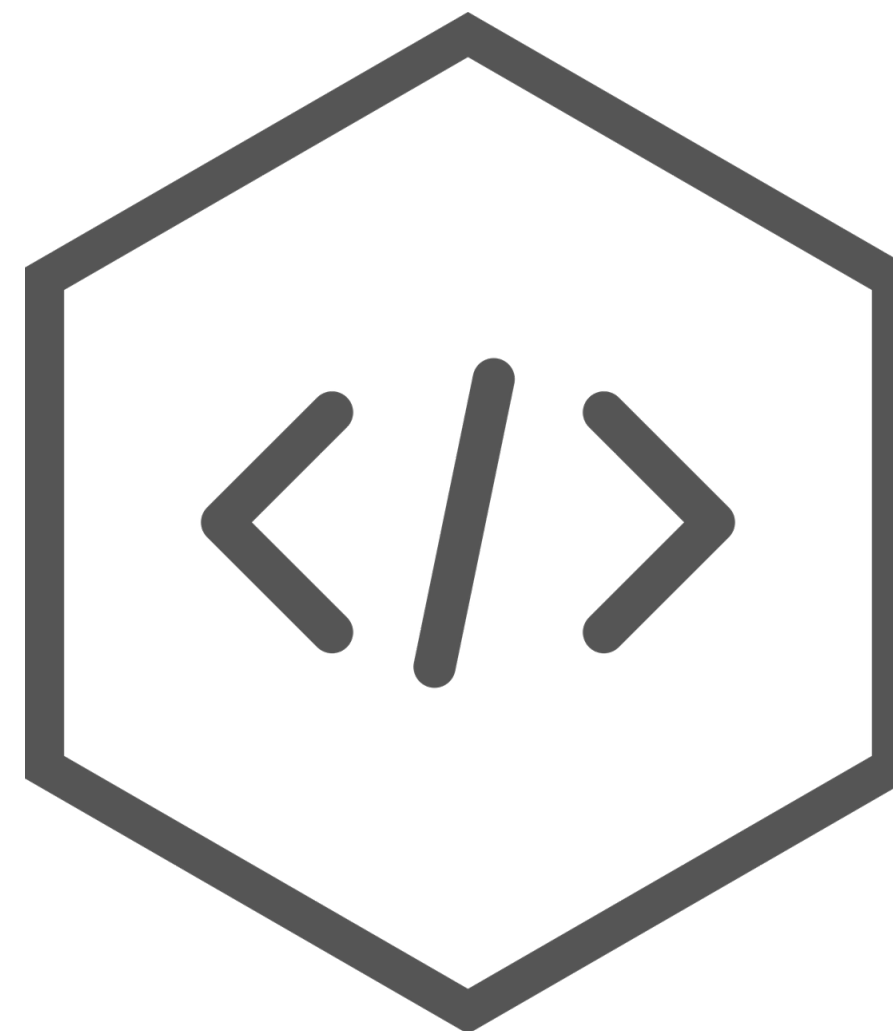


## Introducción

Los documentos que intercambiamos por internet utilizan lenguajes de marcas para expresar el formato del contenido del mismo.

Las **marcas** son señales que delimitan una parte del texto para indicar alguna peculiaridad sobre dicha sección. Los agentes de usuario las interpretan y las aplican.

Las marcas se forman por una palabra que describe su funcionalidad encerrada entre los símbolos menor que (<) y mayor que (>) y es muy **habitual que aparezcan por parejas**, aunque no obligatorio.



### Introducción

**Los lenguajes de marcas no son lenguajes de programación**; sin embargo, se pueden combinar, dentro del mismo documentos con otros que sí lo son como JavaScript o PHP.

Los lenguajes de marcas son **independientes del destinatario final**, de modo que es el intérprete quien representará la información de manera adecuada según nos encontremos en un ordenador, tablet, teléfono móvil, etc.

Con el objetivo de **separar** todavía más la **representación** de una página web de su **contenido**, nació **CSS** (*Cascading Style Sheets*) que no es un lenguaje de marcas sino de estilos.



### SGML. Origen del lenguaje de marcado

Surgió para solucionar los **problemas de compatibilidad** a los que se enfrentaban las empresas de publicación de documentos electrónicos en los años 60.

IBM intentó solucionar los problemas mencionados con el lenguaje de marcas **GML** (*Generalized Markup Language*), que podía ser utilizado por varios dispositivos e incorporaba marcas descriptivas para la estructura de la información.

En 1986 pasó a manos de **ISO** (*International Standardization Organization*) convirtiéndose en **SGML** (ISO 8879), *Standard Generalized Markup Language*.

SGML no es un lenguaje propiamente dicho, sino un metalenguaje que permite definir lenguajes de marcado. Por ejemplo, **HTML es un ejemplo de lenguaje de marcado definido a partir de SGML.**

## SGML. Origen del lenguaje de marcado

Un documento SGML consta de dos partes: **estructura y datos**:



## SGML. Origen del lenguaje de marcado

Estudiando un ejemplo:

- **Declaración:** se ha omitido.
- **DTD** (*Definición de tipo de Documento*): vocabulario y reglas de uso.
- **Instancia:** datos.

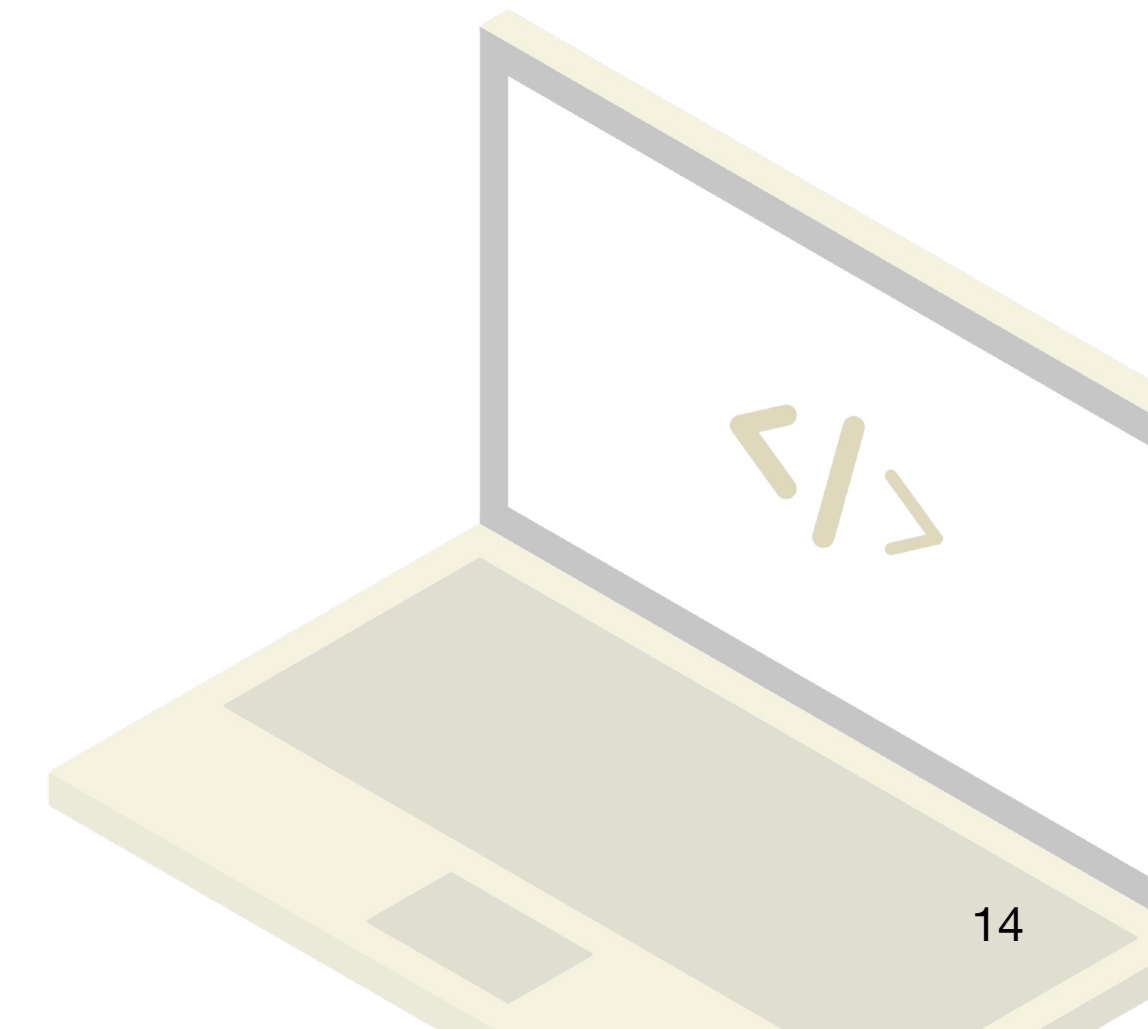
```
<dam>
  <modulo>
    <titulo>Lenguajes de Marcas</titulo>
    <contenido>
      <unidad>Introducción</unidad>
      <unidad>HTML</unidad>
      <unidad>CSS</unidad>
      <unidad>XML</unidad>
      <unidad>XSLT</unidad>
      <unidad>Sindicación</unidad>
      <unidad>ERP</unidad>
    </contenido>
  </modulo>
</dam>
```

### SGML. Origen del lenguaje de marcado

- **Vocabulario:** dam, modulo, titulo, contenido, unidad.
- **Reglas:**
  - DAM contiene varios módulos.
  - Cada módulo contiene un elemento simple llamado titulo y un elemento compuesto denominado contenido.
  - Contenido consta de varias unidades.
  - Toda unidad debe estar en un contenido, todo contenido dentro de un módulo y todos los módulos dentro de DAM.
  - Detrás de cada unidad sólo puede ir otra unidad o fin de contenido.
  - Detrás de cada módulo solo puede ir otro módulo o fin de DAM.

## ACTIVIDAD 1 - Primer HTML

*Busca la estructura básica de un documento HTML y añádele información. Posteriormente, indica su vocabulario y sus reglas.*





### Características de los Lenguajes de Marcas

- **TEXTO PLANO:** Únicamente contienen caracteres de texto, ni imágenes, sonido, archivos comprimidos o cualquier otro tipo propio de los archivos binarios.
- **COMPACIDAD:** Tanto el contenido como las marcas que indican su formato se entremezclan en el fichero componiendo su estructura.
- **INDEPENDENCIA DEL DISPOSITIVO FINAL:** El mismo documento se mostrará de diferentes formas según el intérprete o, dicho de otra forma, el dispositivo final
- **ESPECIALIZACIÓN:** El objetivo que le dio origen fue visualizar documentos de texto; sin embargo, han comenzado a utilizarse en otras áreas como sindicación de contenidos o gráficos vectoriales.
- **FLEXIBILIDAD:** Pueden coordinarse con lenguajes de programación. Por ejemplo, HTML con JavaScript o PHP. Es más, hay etiquetas específicas para ello <script>.



### Clasificación de los Lenguajes de Marcas

#### ATENDIENDO AL TIPO DE MARCAS QUE UTILIZAN

- **DE PRESENTACIÓN:** Indican el formato del texto sin especificar su estructura: tamaño, estilo o justificación de la fuente...

Los procesadores de texto utilizan este tipo de marcado.

Ejemplos: nroff, troff, RTF, TeX, Docbook, ...

- **DESCRIPTIVO, ESTRUCTURAL O SEMÁNTICO:** Especifica las partes en que se divide un documento sin indicar cómo se representará cada una de ellas. XML surge precisamente para generar marcado descriptivo.

Se les denomina bases de datos semiestructuradas porque almacenan información pero no siguen las reglas de integridad propias de las BD relacionales.

- **HÍBRIDO:** Contienen marcas de los dos tipos anteriores: HTML.

## Clasificación de los Lenguajes de Marcas

### ATENDIENDO A SU FUNCIONALIDAD

- **TECNOLOGÍAS DE INTERNET**

- **Páginas Webs:** HTML, XHTML, WML.
- **Formularios/interfaces de usuario:** GladeXML, Xforms, XAML.
- **Sindicación de contenidos:** RSS, Atom.
- **Servicios Web:** WSDL, SOAP, UDDI.
- **Mensajería instantánea:** XMPP

- **DE PROPÓSITO ESPECÍFICO**

- **Gráficos vectoriales:** SVG, X3D, VML
- **Transformación de documentos:** XSLT
- **Partituras de música:** MusicXML
- **Fórmulas matemáticas:** MathML, CML
- ...

### Organizaciones y estándares

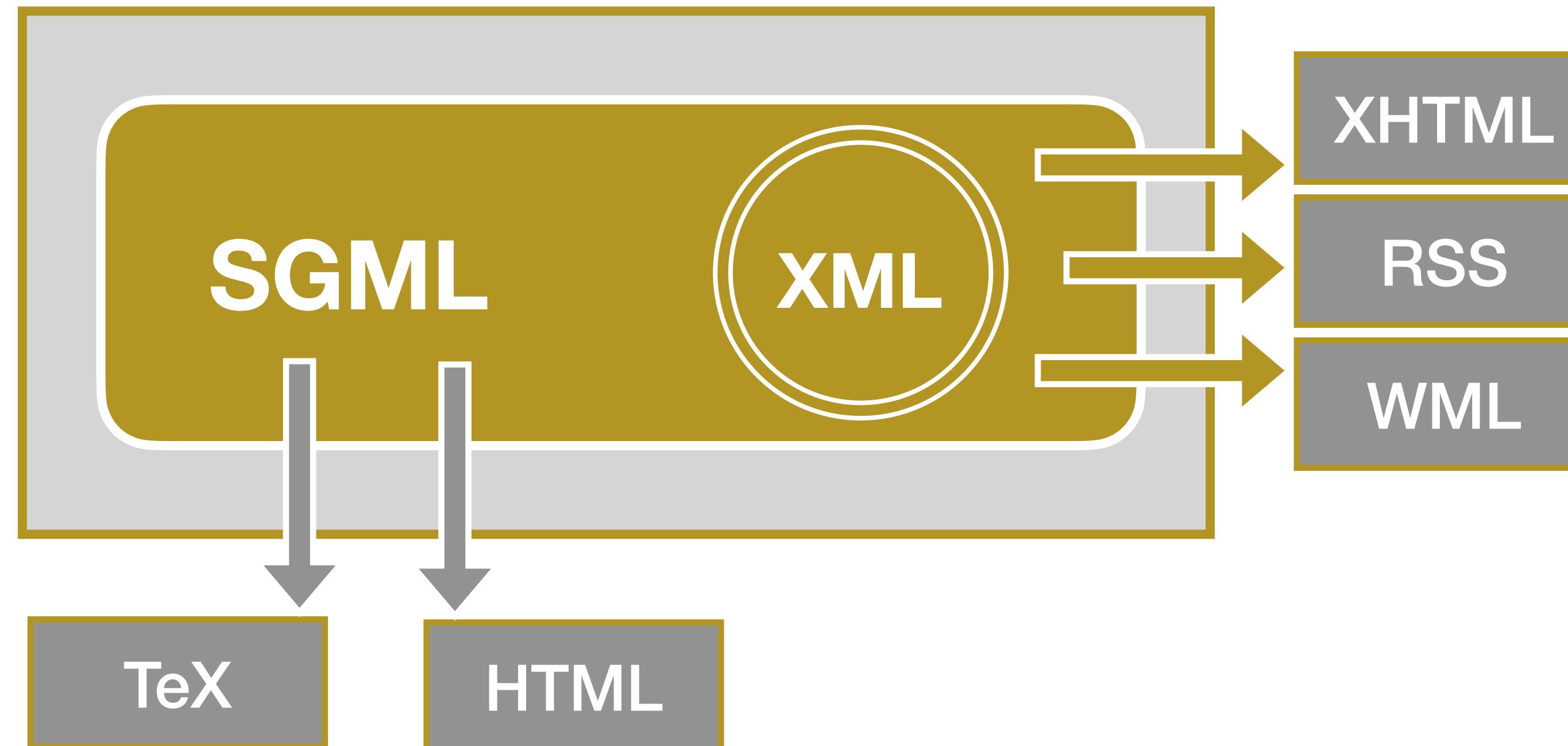
En materia de software destacan organizaciones como W3C, ISO y Open Source.

W3C (World Wide Web Consortium) es una comunidad internacional que desarrolla estándares que aseguran el crecimiento de la Web y que cuenta con miembros como Microsoft, Apple y Google, entre otras.

Además de producir estándares, W3C ha creado software de Código Abierto, entre los que cabe destacar el validador W3C, que utilizaremos durante el curso.



## Principales lenguajes de marcas



### HTML

- Destinado a crear páginas Web.
- Permite mezclar texto y contenido multimedia en un mismo documento.
- Dispone de enlaces o hipervínculos.

### Principales lenguajes de marcas: XML

- Extensible Markup Language.
- Es una simplificación de SGML.
- No impone las etiquetas permitidas y se pueden elegir las necesarias para cada aplicación. Esta característica es la clave de la función principal de XML de representación e intercambio de datos.
  - Posee ventajas significativas
  - Las etiquetas hace que los mensajes sean autodocumentados.
  - El formato no es rígido. Si se agrega información adicional, el remitente puede, simplemente, ignorar la etiqueta. Además, evoluciona con el tiempo y permite representar atributos multivalorados.
  - Permite estructuras anidadas.

### Principales lenguajes de marcas: XML

#### Ejemplo de XML

Los pedidos de compra se generan por una empresa y se envían a otra. Tradicionalmente, el comprador lo imprimía en papel y enviaba al proveedor, que reintroducía manualmente los datos en el sistema informático.

**XML proporciona una forma estándar de etiquetar los datos. Por supuesto, ambas empresas deben estar de acuerdo en qué etiquetas aparecen en el pedido y qué significa cada una de ellas.**



## Principales lenguajes de marcas: XML

```
<pedido_compra>
  <identificador>P-210</identificador>
  <comprador>
    <nombre>José Pérez Salinas</nombre>
    <direccion>Av. Ramón y Cajal, 10</direccion>
    <telefono>678123456</telefono>
  </comprador>
  <proveedor>
    <nombre>Mario Álvarez Vázquez </nombre>
    <direccion>Av. Madrid, 50. Renault</direccion>
    <telefono>983457520</telefono>
  </proveedor>
  <lista_elementos>
    <elemento>
      <identificador>TALISMAN</identificador>
      <cantidad>1</cantidad>
      <descripcion>Llantas aleación 48cm (19") INITIALE PARIS</descripcion>
      <precio>37900</precio>
    </elemento>
  </lista_elementos>
  <coste_total>37900</coste_total>
  <forma_pago>Transferencia</forma_pago>
  <forma_envio>Carretera</forma_envio>
</pedido_compra>
```

### Principales lenguajes de marcas: XML

- El constructor fundamental de XML es el **elemento**.
- Un elemento es un par de etiquetas coincidentes de inicio y finalización, entre las cuales se incluye información.
- Todo documento XML tiene un único **elemento raíz** que contiene a los demás.
- Se dice que un texto aparece en el **contexto de un elemento** si aparece entre la etiqueta de inicio y fin de ese elemento.
- Un elemento que no contiene ni subelementos ni texto, `<elemento></elemento>`, se puede abreviar como `<elemento/>`.



## Principales lenguajes de marcas: XML

### Atributo

Los atributos de un elemento aparecen como pares **nombre=valor** antes del cierre '>' de una etiqueta y únicamente pueden aparecer una vez en una etiqueta dada.

```
<cuenta tipo="corriente">  
  <numero>C-225</numero>  
  <sucursal>Zamora</sucursal>  
  <saldo>13200</saldo>  
</cuenta>
```

### Espacio de nombres

Los espacios de nombres (**namespaces**) son uno de los pilares que conforman los cimientos de XML. Permiten eliminar ambigüedades y solucionar problemas de homonimia.

Surgió la pregunta de **cómo combinar en un mismo documento varios vocabularios.**

```
<pais nombre="España">  
  <capital>Madrid</capital>  
</pais>
```

```
<inversion>  
  <capital>2000€</capital>  
</inversion>
```

### Principales lenguajes de marcas: XML

#### Espacio de nombres

- La solución es anteponer por cada etiqueta o atributo un identificador.
- Los identificadores largos son engorrosos y por ello, se proporciona una forma de definir abreviaturas para ellos.
- Un documento puede tener más de un espacio de nombres y se puede definir uno de ellos como predeterminado.

```
<banco xmlns:BP="http://www.bancoprincipal.com">  
...  
  <BP:sucursal>  
    <BP:nombre>Centro</BP:nombre>  
    <BP:ciudad>Toledo</BP:ciudad>  
  </BP:sucursal>  
...  
</banco>
```

## ACTIVIDAD 2 - Diseño de XML

*Diseña un documento XML que almacene la siguiente información.*

Continente	Ciudad	País	Continente	Ciudad	País
Europa	Berlín	Alemania	África	El Cairo	Egipto
Europa	Viena	Austria	África	Nairobi	Kenia
Europa	Bruselas	Bélgica	África	Rabat	Marruecos
Europa	Copenhague	Dinamarca	África	Dakar	Senegal
América	Buenos Aires	Argentina	Asia	Nueva Delhi	India
América	San José	Costa Rica	Asia	Jerusalén	Israel
América	La Habana	Cuba	Asia	Manila	Filipinas
América	Washington D.C	EEUU	Asia	Bangkok	Tailandia



## ACTIVIDAD 3 - Corrección de errores

Indica los errores que se presentan en el siguiente documento XML

```
<terrestres>
  <vehiculo>coche<vehiculo>
  <vehiculo>tren<vehiculo>
  <vehiculo>bicicleta<vehiculo>
<acuaticos>
  <vehiculo>piragua<vehiculo>
  <vehiculo>catamarán<vehiculo>
  <vehiculo>barco<vehiculo>
<aereos>
  <vehiculo>helicóptero<vehiculo>
  <vehiculo>avión<vehiculo>
```

