LENGUAJES DE MARCAS UD-1

CARACTERÍSTICAS DE LOS LENGUAJES DE MARCAS

ÍNDICE

[INTRODUCCIÓN 3](#_Toc115277946)

[CONCEPTOS 4](#_Toc115277947)

[A. DEFINICIÓN DE LENGUAJE DE MARCAS 4](#_Toc115277948)

[B. EVOLUCIÓN 4](#_Toc115277949)

[C. ORGANISMOS DE ESTANDARIZACIÓN 4](#_Toc115277950)

[D. CARACTERÍSTICAS DE LOS LENGUAJES DE MARCAS 4](#_Toc115277951)

[E. ETIQUETAS 5](#_Toc115277952)

[F. ATRIBUTOS 5](#_Toc115277953)

[G. GRAMÁTICA 6](#_Toc115277954)

[LENGUAJES DE MARCAS EN ENTORNOS WEB 6](#_Toc115277955)

[A. GML (IBM GENERALIZED MARKUP LANGUAGE) 6](#_Toc115277956)

[B. SGML (STARDARD GENERALIZED MARKUP LANGUAGE) 7](#_Toc115277957)

[C. HTML (HYPERTEXT MARKUP LANGUAGE) 7](#_Toc115277958)

[D. XML (EXTENSIBLE MARKUP LANGUAGE) 8](#_Toc115277959)

[E. XHTML (EXTENSIBLE HYPERTEXT MARKUP LANGUAGE) 8](#_Toc115277960)

[CLASIFICACIÓN 8](#_Toc115277961)

[A. MARCAS DE PRESENTACIÓN 8](#_Toc115277962)

[B. MARCAS DE PROCEDIMIENTOS 8](#_Toc115277963)

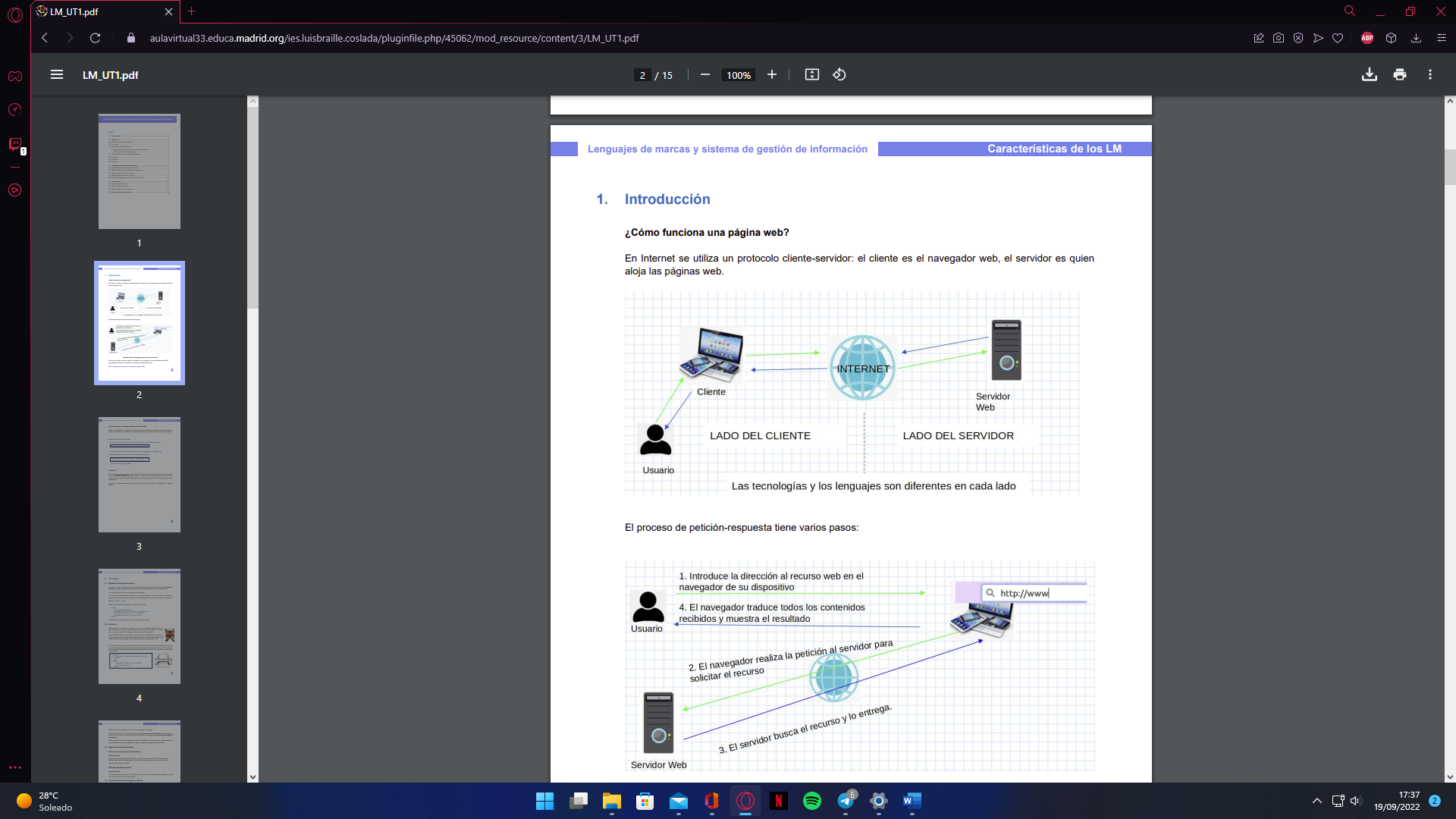
[C. MARCAS DESCRIPTIVAS O SEMÁNTICAS 9](#_Toc115277964)

[ALGUNOS LENGUAJES DE MARCAS 9](#_Toc115277965)

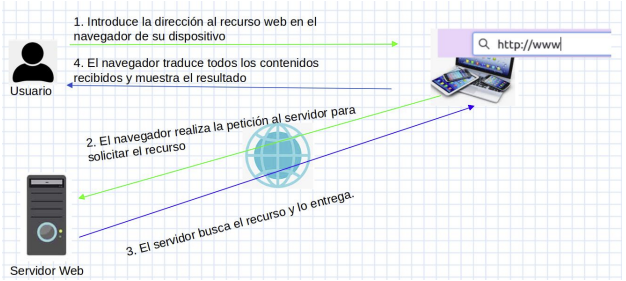
# INTRODUCCIÓN

Los lenguajes de marcas se utilizan en diferentes contextos, uno de ellos es en las páginas web.

En Internet se utiliza un **protocolo cliente-servidor (HTTP)** siendo el cliente el navegador web y no el propio usuario, el servidor es el que aloja las páginas web. Es decir, va a haber un servidor y muchos clientes.

El sistema es una ida-vuelta:

El proceso de petición-respuesta tiene varios pasos:



El lenguaje que utiliza el servidor cuando entrega el recurso al navegador es un lenguaje de marcas denominado HTML el cual se utiliza para indicar el formato de presentación de la información, es decir, cómo queremos visualizarla (no todos los lenguajes de marcas se utilizan con este propósito).

# CONCEPTOS

## DEFINICIÓN DE LENGUAJE DE MARCAS

Es la forma de codificar un documento que incorpora **etiquetas o marcas** que contienen información adicional acerca de la estructura del texto o su presentación.

Un lenguaje de marcas no es un lenguaje de programación.

Las marcas más utilizadas van encerradas entre símbolos de mayor y menor y describen la de principio y la de final. Por ejemplo:

<title> </title>

Un lenguaje de marcas permite **almacenar y formatear información**.

## EVOLUCIÓN

Surgen de manera natural como códigos que los procesadores de texto introducen en los documentos para dirigir el proceso de presentación mediante una impresora. Estaban ligados a las características de la máquina, programa o procesador de textos concretos.

Más tarde se añadieron como medio de presentación a la pantalla.

El marcado al principio estaba orientado a la **presentación** de la información y más tarde a la **forma de estructurar y compartir** la información.

## ORGANISMOS DE ESTANDARIZACIÓN

Los lenguajes son estándares para que las diferentes aplicaciones que los usen se puedan usar. Hay dos estándares:

* **ISO** ***(International Organization for Standarization)*** à PDF
* **W3C** ***(World Wide Web Consortium)*** à HTML, CGI, XML, Xpath, DOM… Recomendaciones y estándares sobre www.

## CARACTERÍSTICAS DE LOS LENGUAJES DE MARCAS

1. Son **archivos de texto plano**. Son texto sin formato, por lo que no requieren ejecución y se pueden editar con un editor de texto sencillo.
2. Son **compactos**. Mezclan en el mismo sitio las etiquetas y el texto.
3. **Facilidad de procesamiento**. No necesitan de compilación, sólo una interpretación.
4. **Independencia del dispositivo**. Al ser un texto estándar, permite crear esos archivos en un sistema y ser compatibles con todos los sistemas.
5. **Flexibilidad**. Son compatibles con otras tecnologías (programas, técnicas, lenguajes de programación…).

## ETIQUETAS

Es una **marca** establecida por el lenguaje.

Los lenguajes las usan de forma **intercaladas en un documento de texto sin formato**. Estas etiquetas son interpretadas a posteriori por los intérpretes y ayudan al procesado del documento.

Dos etiquetas: la de **apertura** y la de **cierre**.

Estructura:

<etiqueta>texto que sufrirá los cambios</etiqueta>

Por ejemplo:

<u>Esto está subrayado</u>

## ATRIBUTOS

Son los **modificadores de etiquetas** que permiten definir las diferentes posibilidades de la etiqueta.

Permiten añadir propiedades a los elementos de un documento. No pueden organizarse jerárquicamente, no pueden contener ningún otro elemento o atributo y no reflejan ninguna estructura lógica.

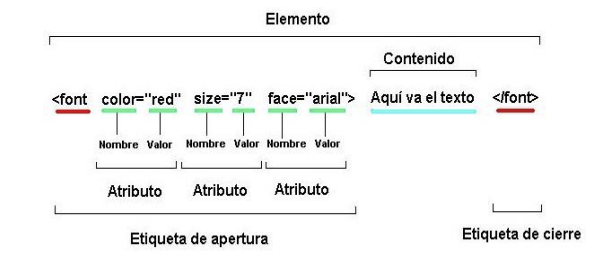
Estructura:

<etiqueta atributo=”valor”>texto afectado</etiqueta>

Ejemplo HTML:

<a href=”web.mit.edu”>Enlace al MIT</a>

Una misma etiqueta puede contener varios atributos:



## GRAMÁTICA

Conjunto de reglas y principios que definen el uso de un lenguaje.

En un lenguaje de marcas la gramática especifica las etiquetas y atributos que se pueden emplear, su orden y sintaxis. Existen varias formas de definir una gramática para un lenguaje de marcas:

* DTD (*Document Type Definition*).
* XSD (*XML Schema Document*).

# LENGUAJES DE MARCAS EN ENTORNOS WEB

## GML (*IBM GENERALIZED MARKUP LANGUAGE*)

Texto

Descripción generada automáticamenteEs el primer lenguaje de marcas, lo creo IBM como respuesta al tamaño que tenía la empresa y la cantidad de información que tenían que crear y compartir. Decidieron crear un **estándar** de documento y formato para que todos los departamentos siguieran las mismas reglas de formato y no crear problemas de compatibilidad con los **diferentes dispositivos de impresión**.

GML añade al texto unas marcas que definen las partes del documento y tienen el formato **:marca** sin etiqueta de cierre.

Evolucionó y se estandarizó como **StandardGML**.

## SGML (*STARDARD GENERALIZED MARKUP LANGUAGE*)

Es estándar ISO. Es un **metalenguaje** (sirve para crear otros lenguajes y utilizarlos con propósito específico).

Es un lenguaje muy complejo y debido a ello no se utiliza actualmente, pero en su momento dio lugar al HTML, al XML y al RTF.

## HTML (*HYPERTEXT MARKUP LANGUAGE*)

Es el lenguaje para **hacer páginas web**, es una simplificación de SGML. Existen multitud de herramientas software para manipular e interpretar el lenguaje HTML.

La visualización del formato de un documento HTML se realiza a través de un navegador web.

Tiene varias **limitaciones**:

* **No soporta tareas de impresión y diseño**.
* **Etiquetas limitadas**.
* En su inicio se creó sin tener en cuenta los diferentes usos que se le podían dar, eso ha provocado la **generación de diferentes estándares** y algunos navegadores crearon sus propias etiquetas no estandarizadas. Además, algunos navegadores pueden tener diferencias en la interpretación de una misma etiqueta.
* **No permite mostrar contenido dinámico**. Se complementa con tecnologías como JavaScript.
* **El contenido y el formato están mezclados en el documento**. Esto no ocurre si se utiliza CSS.

HTML es un **estándar compuesto por recomendaciones publicadas por el W3C**. Existen varias versiones del lenguaje.

La primera etiqueta de un documento HTML hace referencia a la versión que se utiliza:

Texto

Descripción generada automáticamente

## XML (EXTENSIBLE MARKUP LANGUAGE)

Es una simplificación y adaptación de SGML que permite definir una gramática para lenguajes específicos, es decir, es un metalenguaje.

Está publicada por W3C.

Sus características son:

* Es un **metalenguaje**.
* **No incluye ninguna información relativa al formato**.

La forma de XML es muy parecida a la de un documento HTML (etiquetas).

En un documento XML la primera etiqueta hace referencia a la versión que se utiliza:



## XHTML (EXTENSIBLE HYPERTEXT MARKUP LANGUAGE)

Es HTML al que se le aplica un conjunto de normas derivadas del XML. Es básicamente HTML expresado como XML válido.

Es más estricto a nivel técnico, pero esto permite que posteriormente sea más sencillo hacer cambios o buscar errores.

# CLASIFICACIÓN

## MARCAS DE PRESENTACIÓN

Son lenguajes cuyo propósito es presentar la información y dar formato al texto:

* Cada marca tiene apertura y cierre.
* Los documentos son fáciles de procesar, pero es difícil mantenerlos.
* Las marcas se suelen ocultar al usuario.

Por ejemplo: Lenguaje RTF.

## MARCAS DE PROCEDIMIENTOS

Son lenguajes orientados a la presentación de la información, pero en este caso las etiquetas indican los procedimientos que debe seguir el software encargado de la presentación. Por ejemplo: Tex, LaTex.

## MARCAS DESCRIPTIVAS O SEMÁNTICAS

Son lenguajes que pretenden marcar el texto y definir qué es cada parte del texto. Da información semántica. Por ejemplo: Lenguaje SGML, XML.

# ALGUNOS LENGUAJES DE MARCAS

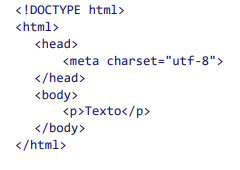
Existen diferentes lenguajes de marcas, por ejemplo:

* **RTF (*Rich Text Format*)**: Permite el intercambio de documentos de texto entre distintos procesadores de texto.
* **TeX**: Permite crear ecuaciones matemáticas complejas.

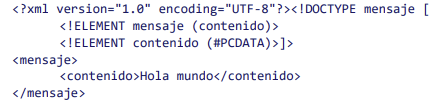
* **Wikitexto**: Permite la creación de páginas wiki.

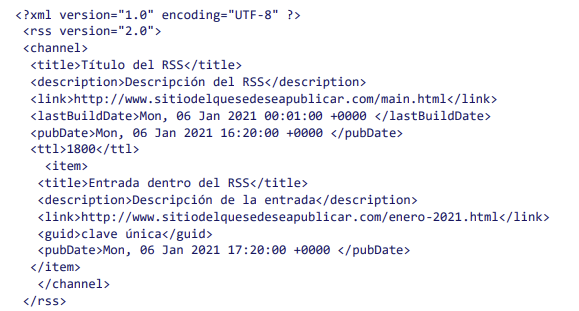


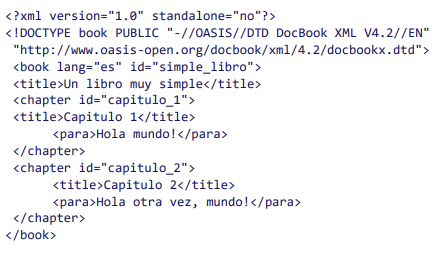
* **HTML, XHTML**: Permite la creación de páginas web.



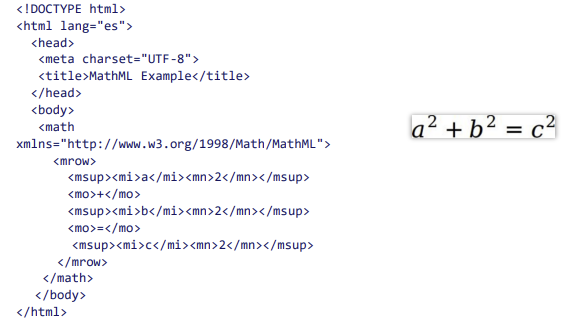
* **XML**: Estructura datos para facilitar su intercambio y almacenaje.

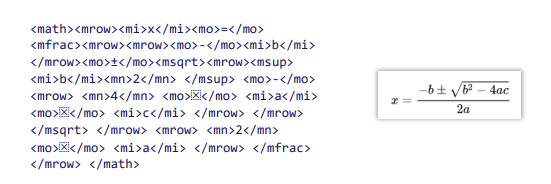


* **RSS**: Se utiliza para difundir información actualizada frecuentemente a usuarios que se han suscrito a una fuente de contenidos. Basado en XML.
* **DocBook**: Se utiliza para estructurar documentos, libros, manuales… Es una aplicación de XML con DTD propia.

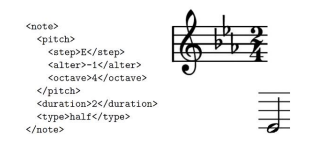


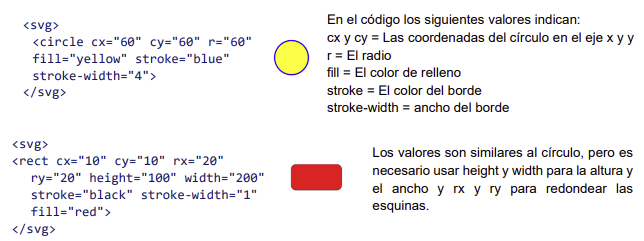
* **MathML**: Basado en XML y se utiliza para la expresión de la notación matemática. Actualmente no sólo se utiliza en las páginas web, se puede usar por sí mismo.





* **MusicXML**: Permite el intercambio de partituras entre distintos editores de partituras.



* **SVG**: Basado en XML, sirve para la representación de gráficos vectoriales (son aquellos que tienen escalado constante impidiendo que se pixelen).

