

# TEMA 01

## HTML: Conceptos Básicos

# Lenguajes de marcas

- Un lenguaje de marcas es aquel que combina la información con un conjunto de elementos que indican cómo representarla.
- Estos elementos son utilizados por un intérprete para presentar la información y no se representan en el resultado final.
- **Tipos de lenguajes de marcas:**
  - Orientados a presentación:
    - Ocultan al usuario los elementos
    - Ejemplo: Microsoft Word

# Lenguajes de marcas

- **Tipos de lenguajes de marcas:**

- Procedurales:

- Están orientados a la presentación pero además utilizan acciones y procedimientos
- Ejemplo: Latex, Postscript

- Descriptivos:

- Describen qué es cada elemento para que un intérprete los represente como considere
- Ejemplo: SGML, HTML, XML, XHTML

# SGML

- **Standard Generalized Markup Language**
  - Lenguaje normalizado por la ISO
  - Establece las normas para etiquetar documentos
  - A finales de los 80, basándose en SGML, se crea un lenguaje de marcas orientado a documentos que se transmiten por la red. Nace **HTML**.

```
</TEIHEADER>
<TEXT>      <T Tit><L 1><Y Q><P cc4><C B>
1      <Z {Actus Primus. Scoena Prima}.>
2      *D {Flourish. Enter the Tribunes and Senators aloft And then}
3      {enter Saturninus and his Followers at one doore}.
4      {and Bassianus and his Followers at the}
5      {other. with Drum & Colours}.>
6      <S {Saturninus}.>
7      Noble Patricians, Patrons of my right,
8      Defend the iustice of my Cause with Armes.
9      And Countrey- men, my louing Followers,
10     *Pleade my Successiue Title with your Swords.
11     I was the first borne Sonne, that was the last
12     That wore the Imperiall Diadem of Rome:
13     Then let my Fathers Honours liue in me,
14     Nor wrong mine Age with this indignitie.
15     <S {Bassianus}.> Romaines, Friends, Followers,
16     Fauourers of my Right:
17     If euer {Bassianus}, {Caesars} Sonne,
18     Were gracious in the eyes of Royall Rome,
```

- **HyperText Markup Language**

- Fue ideado por Tim Berners-Lee en 1980
- 1993 se intentó crear un estándar pero no se consiguió
- HTML se siguió desarrollando pasando a su versión 2.0, hasta que en 1995 se consiguió su estandarización
- El primer estándar oficial de HTML es la versión 2.0
- A partir de 1996, es la W3C (World Wide Web Consortium), quien se encarga de los estándares de HTML
- En enero de 1997 publican su primer estándar: HTML 3.2
- En abril de 1998 publican HTML 4.0, que revoluciona notablemente todo el diseño web
- 16 años después, en octubre de 2014, la W3C publica por fin el estándar para HTML5.

- **Extensible Markup Language**
  - Fue desarrollado por el W3C
  - También está basado en SGML
  - Es un estándar para definir la gramática de lenguajes de marcas
  - Nace por la poca rigurosidad gramatical que existía en algunos lenguajes como HTML

# XHTML

- **Extensible HyperText Markup Language**
  - Lenguaje desarrollado por el W3C basado en HTML pero con la sintaxis de XML
  - Aunque era técnicamente mejor que HTML, el aumentar la complejidad, ha hecho que no haya sido muy bien aceptado

# HTML5

- **Versión 5 de HTML.**
  - Cuando W3C se centra en XHTML deja de lado a HTML
  - Un conjunto de empresas, al ver la necesidad de mejorar HTML sin que W3C haga nada, deciden crear una asociación, la WHATWG (*Web Hypertext Application Technology Working Group*)
  - La WHATWG se centra en HTML5 hasta que en 2007 la W3C, por la presión de las empresas y la WHATWG, decide retomar el proyecto
  - En 2008 se publica el primer borrador de HTML5 y en ese mismo momento diversos navegadores comienzan sus tareas de adaptación





# HTML5

- En 2009, la W3C renuncia a su desarrollo de XHTML 2.0 y reconoce a HTML5 como la próxima generación del lenguaje
- Comienzan a trabajar en paralelo, pero en 2012 WHATWG opta por un “living standard” de HTML5 (estándar sin fin con continua evolución) mientras que W3C opta por un estándar único con fecha de publicación en 2014
- La última versión publicada es la 5.3 del 18 de octubre de 2018

# Revisión de Páginas

- La W3C ofrece dos páginas para hacer una revisión de la sintaxis de una página web.
- Una web sin errores, puede alojar el logotipo de la W3C
  - <http://validator.w3.org>



The screenshot shows the W3C Markup Validation Service interface. At the top, it says "W3C Markup Validation Service" and "Check the markup (HTML, XHTML, ...) of Web documents". Below this, there's a "Jump To:" section with "Validation Output" selected. A red banner indicates "Errors found while checking this document as XHTML 1.0 Strict!". The main section shows the validation results: "Result: 119 Errors, 62 warning(s)". Below this, there are fields for "Address" (http://www.fcbarcelona.es/), "Encoding" (utf-8), "Doctype" (XHTML 1.0 Strict), "Root Element" (html), and "Root Namespace" (http://www.w3.org/1999/xhtml). There's a "I ♥ VALIDATOR" button and a "Donate" link. The "Options" section includes checkboxes for "Show Source", "Validate error pages", "Show Outline", "Verbose Output", "List Messages Sequentially", "Group Error Messages by Type", and "Clean up Markup with HTML-Tidy". A "Revalidate" button is at the bottom right. The "Validation Output" section shows "119 Errors" and a specific error message: "Line 4, Column 64: syntax of attribute value does not conform to declared value".

# Lecturas Recomendadas

- WHATWG
  - <http://www.whatwg.org>
- W3C
  - <http://www.w3.org>
- Documentación de HTML en inglés
  - <http://www.w3.org/TR/html401/>
- Documentación de HTML en español
  - <http://html.conclase.net/w3c/html401-es/cover.html>
- W3Schools
  - <http://www.w3schools.com>

# Componentes de un lenguaje de marcas

- Existen tres tipos de componentes para componer un lenguaje de marcas:
  - **Etiquetas (*tags*) :**
    - Son textos que van encerrados entre los símbolos (<) y (>)
    - Existen etiquetas de **inicio** (<*tag*>) y de **fin** (</*tag*>)
    - No existe una norma para escribir las etiquetas con mayúsculas o minúsculas. Los navegadores las entienden igualmente
    - En las primeras versiones de HTML se escribían con mayúsculas
    - Los lenguajes de programación (HTML no lo es) se escriben con minúsculas
    - Cada vez más gente las escribe con minúsculas por ser una escritura más rápida y más cómoda

# Componentes de un lenguaje de marcas

- **Elementos:**
  - Constituyen las **partes que componen un documento**
  - Generalmente están formados por:
    - Etiqueta de inicio
    - Contenido
    - Etiqueta de fin
  - Algunos elementos no tienen **contenido** y se les denomina **elementos vacíos** y **no es obligatorio incluir su etiqueta final**

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
<html>
  <head>
  </head>

  <body>
    <p> Esto es un texto
  </body>
</html>
```

# Componentes de un lenguaje de marcas

- **Atributos:**

- Se utilizan para especificar **ciertas propiedades** de los elementos
- Se incluyen dentro de la **etiqueta de inicio**
- Su **sintaxis** es: *atributo=valor*
- Si el valor contiene espacios, comillas, apóstrofes o algún signo especial, **es obligatorio utilizar comillas**
- Existen atributos con **valor booleano**. Que se incluyan o no es lo que indica su valor verdadero o falso
- Ejemplo:

```
<html>
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=utf-8">
  <title>Desarrollo web</title>
</head>

<body bgcolor="#999999">
  Desarrollo web
</body>
</html>
```

- **Definición de Tipo de Documento (DTD):**
  - Contiene las **normas para crear un documento**, es decir, qué elementos son necesarios, cómo utilizarlos, etc.
  - La W3C ofrece tres DTD para trabajar con HTML:
    - ***Strict***: es la DTD recomendada. No reconoce elementos y atributos obsoletos o desaprobados. Está destinada a la separación de presentación y contenido
    - ***Transitional***: incluye lo mismo que la DTD *Strict* más todos los elementos obsoletos
    - ***Frameset***: incluye lo mismo que la DTD *Transitional* más todo lo relacionado con documentos que incluyen marcos, pero los marcos ya están obsoletos

# DTD

- **Definición de Tipo de Documento (DTD):**
  - La forma de incluir un DTD es:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"  
    "http://www.w3.org/TR/html4/strict.dtd">
```

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"  
    "http://www.w3.org/TR/html4/loose.dtd">
```

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN"  
    "http://www.w3.org/TR/html4/frameset.dtd">
```



# DTD en HTML5

- En **HTML5** el **DOCTYPE** se escribe

```
<!DOCTYPE html>
```

