

TEMA 1: LENGUAJE HTML

1. ¿Qué es HTML?

Definiéndolo de forma sencilla, *"HTML es lo que se utiliza para crear todas las páginas web de Internet"*. Más concretamente, HTML es el *lenguaje* con el que se *"escriben"* la mayoría de páginas web.

Los diseñadores utilizan el lenguaje HTML para crear sus páginas web, los programas que utilizan los diseñadores generan páginas escritas en HTML y los navegadores que utilizamos los usuarios muestran las páginas web después de leer su contenido HTML.

Aunque HTML es un lenguaje que utilizan los ordenadores y los programas de diseño, es muy fácil de aprender y escribir por parte de las personas. En realidad, HTML son las siglas de *HyperText Markup Language*.

El lenguaje HTML es un estándar reconocido en todo el mundo y cuyas normas define un organismo sin ánimo de lucro llamado World Wide Web Consortium, más conocido como **W3C**. Como se trata de un estándar reconocido por todas las empresas relacionadas con el mundo de Internet, una misma página HTML se visualiza de forma muy similar en cualquier navegador de cualquier sistema operativo. El propio **W3C** define el lenguaje HTML como *"un lenguaje reconocido universalmente y que permite publicar información de forma global"*. Desde su creación, el lenguaje HTML ha pasado de ser un lenguaje utilizado exclusivamente para crear documentos electrónicos a ser un lenguaje que se utiliza en muchas aplicaciones electrónicas como buscadores, tiendas online y banca electrónica.

2. Versiones de HTML

El origen de HTML se remonta a 1980, cuando el físico **Tim Berners-Lee**, trabajador del CERN (*Organización Europea para la Investigación Nuclear*) propuso un nuevo sistema de *"hipertexto"* para compartir documentos.

Los sistemas de *"hipertexto"* habían sido desarrollados años antes. En el ámbito de la informática, el *"hipertexto"* permitía que los usuarios accedieran a la información relacionada con los documentos electrónicos que estaban visualizando. De cierta manera, los primitivos sistemas de *"hipertexto"* podrían asimilarse a los enlaces de las páginas web actuales.

Tras finalizar el desarrollo de su sistema de *"hipertexto"*, Tim Berners-Lee lo presentó a una convocatoria organizada para desarrollar un sistema de *"hipertexto"* para Internet. Después de unir sus fuerzas con el ingeniero de sistemas **Robert Cailliau**, presentaron la propuesta ganadora llamada *WorldWideWeb (W3)*.

El primer documento formal con la descripción de HTML se publicó en 1991 bajo el nombre *"HTML Tags" (Etiquetas HTML)*.

La primera propuesta oficial para convertir HTML en un estándar se realizó en 1993 por parte del organismo IETF (*Internet Engineering Task Force*). Aunque se consiguieron

avances significativos ninguna de las dos propuestas de estándar, llamadas HTML y HTML+ consiguieron convertirse en estándar oficial.

En 1995, el organismo IETF organiza un grupo de trabajo de HTML y consigue publicar, el estándar HTML 2.0. A pesar de su nombre, HTML 2.0 es el primer estándar oficial de HTML.

A partir de 1996, los estándares de HTML los publica otro organismo de estandarización llamado W3C (*World Wide Web Consortium*).

HTML 4.0 se publicó en Abril de 1998 y supone un gran salto desde las versiones anteriores. Entre sus novedades más destacadas se encuentran las hojas de estilos CSS, la posibilidad de incluir pequeños programas o *scripts* en las páginas web, mejora de la accesibilidad de las páginas diseñadas, tablas complejas y mejoras en los formularios.

La última especificación oficial de HTML se publicó en diciembre de 1999 y se denomina HTML 4.01. Se trata de una revisión y actualización de la versión HTML 4.0, por lo que no incluye novedades significativas.

Desde la publicación de HTML 4.01, la actividad de estandarización de HTML se detuvo y el W3C se centró en el desarrollo del estándar XHTML. Por este motivo, en el año 2004, las empresas Apple, Mozilla y Opera mostraron su preocupación por la falta de interés del W3C en HTML y decidieron organizarse en una nueva asociación llamada WHATWG (*Web Hypertext Application Technology Working Group*).

La actividad actual del WHATWG se centra en el estándar HTML 5, cuyo primer borrador oficial se publicó en enero de 2008. Debido a la fuerza de las empresas que forman el grupo WHATWG y a la publicación de los borradores de HTML 5.0, en marzo de 2007 el W3C decidió retomar la actividad estandarizadora de HTML.

3. ¿Por qué se dice que HTML es un lenguaje de marcado?

Uno de los retos iniciales a los que se tuvo que enfrentar la informática fue el de cómo almacenar la información en los archivos digitales. Como los primeros archivos sólo contenían texto sin formato, la solución utilizada era muy sencilla: se codificaban las letras del alfabeto y se transformaban en números. De esta forma, para almacenar un contenido de texto en un archivo electrónico, se utiliza una tabla de conversión que transforma cada carácter en un número. Una vez almacenada la secuencia de números, el contenido del archivo se puede recuperar realizando el proceso inverso.

El proceso de transformación de caracteres en secuencias de números se denomina **codificación de caracteres** y cada una de las tablas que se han definido para realizar la transformación se conocen con el nombre de **páginas de código**. Una de las codificaciones más conocidas es la codificación ASCII.

Una vez resuelto el problema de almacenar el texto simple, se presenta el reto de almacenar los contenidos de texto con formato. En otras palabras, ¿cómo se almacena un texto en negrita? ¿y un texto de color rojo? ¿y otro texto azul, en negrita y subrayado?

Utilizar una tabla de conversión similar a las que se utilizan para textos sin formato no es posible, ya que existen infinitos posibles estilos para aplicar al texto. Una solución técnicamente viable consiste en almacenar la información sobre el formato del texto en

una zona especial reservada dentro del propio archivo. En esta zona se podría indicar dónde comienza y dónde termina cada formato. No obstante, la solución que realmente se emplea para guardar la información con formato es mucho más sencilla: el archivo electrónico almacena tanto los contenidos como la información sobre el formato de esos contenidos. Si por ejemplo se quiere dividir el texto en párrafos y se desea dar especial importancia a algunas palabras, se podría indicar de la siguiente manera:

<parrafo>

Contenido de texto con <importante>algunas palabras</importante> resaltadas de forma especial.

</parrafo>

El proceso de indicar las diferentes partes que componen la información se denomina **marcar** (*markup* en inglés). Cada una de las palabras que se emplean para marcar el inicio y el final de una sección se denominan **etiquetas**.

Aunque existen algunas excepciones, en general las etiquetas se indican por pares y se forman de la siguiente manera:

- Etiqueta de apertura: carácter <, seguido del nombre de la etiqueta y terminado con el carácter >
- Etiqueta de cierre: carácter <, seguido del carácter /, seguido del nombre de la etiqueta y terminado con el carácter >

HTML es un **lenguaje de etiquetas** (también llamado **lenguaje de marcado**) y las páginas web habituales están formadas por cientos o miles de pares de etiquetas. De hecho, las letras "ML" de la sigla HTML significan "*markup language*", que es como se denominan en inglés a los *lenguajes de marcado*.

La principal ventaja de los lenguajes de etiquetas es que son muy sencillos de leer y escribir por parte de las personas y de los sistemas electrónicos. La principal desventaja es que pueden aumentar mucho el tamaño del documento, por lo que en general se utilizan etiquetas con nombres muy cortos.

4. Estructura de una página

Las páginas HTML se dividen en dos partes: la cabecera y el cuerpo. La cabecera incluye información sobre la propia página, como por ejemplo su título y su idioma. El cuerpo de la página incluye todos sus contenidos, como párrafos de texto e imágenes.

El cuerpo (llamado *body* en inglés) contiene todo lo que el usuario ve en su pantalla y la cabecera (llamada *head* en inglés) contiene todo lo que no se ve (con la única excepción del título de la página, que los navegadores muestran como título de sus ventanas).

Es importante conocer las tres etiquetas principales de un documento HTML (<html>, <head>, <body>):

- <html>: indica el comienzo y el final de un documento HTML. Ninguna etiqueta o contenido puede colocarse antes o después de la etiqueta <html>. En el interior de la etiqueta <html> se definen la cabecera y el cuerpo del documento HTML y todo lo que se coloque fuera de la etiqueta <html> se ignora.

- <head>: delimita la parte de la cabecera del documento. La cabecera contiene

información sobre el propio documento HTML, como por ejemplo su título y el idioma de la página. Los contenidos indicados en la cabecera no son visibles para el usuario, con la excepción de la etiqueta <title>, que se utiliza para indicar el título del documento y que los navegadores lo visualizan en la parte superior izquierda de la ventana del navegador.

- <body>: delimita el cuerpo del documento HTML. El cuerpo encierra todos los contenidos que se muestran al usuario (párrafos de texto, imágenes, tablas).

A continuación se muestra el código HTML de una página web muy sencilla:

```
<html>

<head>
<title>El primer documento HTML</title>
</head>

<body>
<p>El lenguaje HTML es <strong>tan sencillo</strong> que
prácticamente se entiende sin estudiar el significado
de sus etiquetas principales.</p>
</body>

</html>
```

Es vital que al guardar un archivo en el que figure el código de una página escrita en HTML, se guarde con extensión .htm o .html, ya que solo por la extensión reconocerá el sistema operativo que se trata de un documento web y no de un simple archivo de texto.

El nombre del archivo deberá escribirse en minúsculas y sin espacios ni caracteres especiales. Los únicos signos de puntuación admitidos son el punto (.), que se usará para separar el nombre de la extensión, y el guión bajo. El nombre podrá contener letras y números, pero deberá empezar por una letra. Así mismo nos abstendremos de meter en el nombre de un archivo letras acentuadas, eñes, cedillas o cualquier otro carácter que no pertenezca al alfabeto internacional (inglés).

Aunque estas precauciones parecen excesivas ahora, es de tener en cuenta que los servidores web son muy sensibles a determinados aspectos del nombre de los ficheros. Otra norma importante es que las carpetas en las que se guarden los archivos, no puede aparecer ningún punto en el nombre que se le asigne.

Otra cosa más, HTML no es sensible al uso de mayúsculas y minúsculas. Funciona igual si tecleamos los nombres de las etiquetas y atributos en mayúsculas, minúsculas o combinación de ambos. Como regla general, nos acostumbraremos a escribirlos en minúsculas.

5. Configurando el fondo

Son varias las cosas que podemos configurar del fondo. Por ejemplo, ¿te has fijado en la barra de desplazamiento que aparece a la derecha del área de navegación? Bueno, pues resulta muy útil si la página es lo suficientemente grande como para no caber completa en la pantalla. De esta forma puede desplazarse arriba y abajo. Pero si la

página cabe en el área de navegación sin necesidad de desplazarse, la barra no es necesaria. Y, aunque está inhabilitada, sigue apareciendo. Para ello tenemos el atributo **scroll** que admite los valores “no” (si no queremos que aparezca la barra) o “auto”(si queremos que aparezca solo cuando sea necesario).

```
<body scroll="no">  
<body scroll="auto">
```

Ahora vamos a tratar de alinear el texto en la página. En primer lugar, el texto se coloca automáticamente en el lado izquierdo de la página, salvo un pequeño margen. Para modificar o eliminar ese margen, se necesitan los siguientes atributos:

```
<body leftmargin=0 topmargin=0 rightmargin=0>
```

Los atributos anteriores modifican el margen izquierdo, superior y derecho, respectivamente. Los márgenes se especifican en número de píxeles.

Quizás, una de las posibilidades más llamativas a la hora de montar una página sea el uso de **colores** para el **fondo** de la misma **y** también, el uso de colores para el **texto**. Pues bien, para ello se utilizan los siguientes atributos dentro de la etiqueta del body:

```
<body bgcolor="yellow">  
<body bgcolor="#FFFF00">  
<body text="#0000FF">  
<body bgcolor="blue" text="#FFFF00">
```

Como se puede observar en los ejemplos anteriores, hay dos formas de indicar los colores: con el nombre en inglés o mediante un código en hexadecimal.

6. Colores en HTML

Podemos establecer los colores de fondo y texto con su nombre en inglés, por ejemplo: yellow, red, pink, blue, etc. Esto es válido en la mayoría de los casos. De hecho, existe una variedad suficiente de nombres de colores como para satisfacer todas nuestras necesidades.

Pero existe una alternativa de establecer el color. Para ello, necesitamos saber cómo se crean los colores en el monitor. Para definir un color cualquiera en el monitor se hace a partir de una combinación de tres colores: rojo, verde y azul. Todos los colores que puedas ver en tu monitor están formados a partir de la mezcla de estos tres colores en diferentes proporciones.

Cualquier color que necesitemos, se lo podemos indicar a nuestro código HTML mediante una expresión en la que figurarán cada uno de los colores primarios con un valor entre 0 y 255. Por ejemplo, el amarillo está compuesto por el máximo de rojo y verde y nada de azul, es decir, sería la combinación 255,255,0.

Los colores primarios se expresan siempre en este orden: primero la proporción de rojo, después la de verde y por último, la de azul.

Pero no todo es tan fácil, pues los creadores de HTML decidieron que los valores de los tres colores primarios se expresaran en hexadecimal y con dos dígitos.

Como sabes, la numeración hexadecimal utiliza los dígitos del 0 al 9 y las letras de la A a la F. Por tanto, el 0 sería 00 hexadecimal; el 15 decimal sería 0F hexadecimal; el 255 decimal sería FF; y así sucesivamente.

Como indicamos anteriormente en el ejemplo del color amarillo, en hexadecimal se correspondería con FFFF00.

Una cosa más acerca de los colores. Existen lo que se llaman colores seguros para la web. Son colores seguros para la web aquellos que se visualizan exactamente igual en una plataforma PC que en un Macintosh, ya sea con un navegador Internet Explorer o Netscape Navigator. Una vez más, tenemos que tener en cuenta que estamos trabajando para la res y que nuestra página puede ser visitada por cualquier usuario con distintas plataformas. Los colores seguros para la web son aquellos que están formados por cualquier combinación de los valores 00, 33, 66, 99, CC o FF.

También hay que tener en cuenta que el color del texto debe contrastar con el del fondo, de tal forma que la página sea cómoda para el usuario. En general, se optará por tonos pastel o apagados para el fondo. Los colores chillones se emplean casi exclusivamente para páginas de juegos, música moderna o eróticas.

Tabla de colores en HTML:

DarkTurquoise 00CED1	MediumTurquoise 48D1CC	Turquoise 40E0D0	Cyan 00FFFF	Snow FFFAFA	GhostWhite F8F8FF	WhiteSmoke F5F5F5	Gainsboro DCDCDC
LightCyan E0FFFF	CadetBlue 5F9EAD	MediumAquamarine 66CDAA	Aquamarine 7FFFD4	FloralWhite FFFAF0	OldLace FDF5E6	Linen FAF0E6	AntiqueWhite FAEBD7
DarkGreen 006400	DarkOliveGreen 556B2F	DarkSeaGreen 8FBC8F	SeaGreen 2E8B57	PapayaWhip FFEDD5	BlanchedAlmond FFEBCD	Bisque FFFAF0	PeachPuff FFDAB9
MediumSeaGreen 3CB371	LightSeaGreen 20B2AA	PaleGreen 98FB98	SpringGreen 00FF7F	NavajoWhite FFDEAD	Moccasin FFE4B5	Cornsilk FFF8DC	Ivory FFFFF0
LawnGreen 7CFC00	Green 00FF00	Chartreuse 7FFF00	MedSpringGreen 00FA9A	LemonChiffon FFFACD	Seashell FFF5EE	Honeydew F0FFF0	MintCream F5FFFA
GreenYellow ADFF2F	LimeGreen 32CD32	YellowGreen 9ACD32	ForestGreen 228B22	Azure F0FFFF	AliceBlue F0F8FF	lavender E6E6FA	LavenderBlush FFF0F5
OliveDrab 6B8E23	DarkKhaki BDB76B	PaleGoldenrod EEE8AA	LtGoldenrodYello FAFAD2	MistyRose FFE4E1	White FFFFFF	Black 000000	DarkSlateGray 2F4F4F
LightYellow FFFFE0	Yellow FFFF00	Gold FFD700	LightGoldenrod EEDD82	LightSlateGray 778899	Grey BEBEBE	LightGray D3D3D3	MidnightBlue 191970
Goldenrod DAA520	DarkGoldenrod B8860B	RosyBrown BC8F8F	IndianRed CD5C5C	CornflowerBlue 6495ED	DarkSlateBlue 483D8B	SlateBlue 6A5ACD	MediumSlateBlue 7B68EE
SaddleBrown 8B4513	Sienna A0522D	Peru CD853F	Burlywood DEB887	LightSlateBlue 8470FF	NavyBlue 000080	DimGrey 696969	SlateGrey 708090
Beige F5F5DC	Wheat F5DEB3	SandyBrown F4A460	Tan D2B48C	MediumBlue 0000CD	RoyalBlue 4169E1	Blue 0000FF	DodgerBlue 1E90FF
Chocolate D2691E	Firebrick B22222	Brown A52A2A	DarkSalmon E9967A	DeepSkyBlue 00BFFF	SkyBlue 87CEEB	LightSkyBlue 87CEFA	SteelBlue 4682B4
Salmon FA8072	LightSalmon FFA07A	Orange FFA500	DarkOrange FF8C00	LightSteelBlue B0C4DE	LightBlue ADD8E6	PowderBlue B0E0E6	PaleTurquoise AFEEEE
		Coral FF7F50	LightCoral F08080	Tomato FF6347	OrangeRed FF4500		
		Red FF0000	HotPink FF69B4	DeepPink FF1493	Pink FFC0CB		
		MediumVioletRed C71585	VioletRed D02090	Magenta FF00FF	Violet EE82EE		
		Plum DDA0DD	LightPink FFB6C1	PaleVioletRed DB7093	Maroon B03060		
		Orchid DA70D6	MediumOrchid BA55D3	DarkOrchid 9932CC	DarkViolet 9400D3		
		BlueViolet 8A2BE2	Purple A020F0	MediumPurple 9370DB	Thistle D8BFD8		