

1. Del HTML al XHTML: evolución y versiones

HTML es el lenguaje utilizado para crear la mayor parte de las páginas web. Es un estándar reconocido en todos los navegadores, por lo tanto, todos ellos visualizan una página **HTML** de forma muy similar independientemente del sistema operativo sobre el que se ejecutan.

La evolución de sus versiones, desde su creación hasta la creación del **XHTML** podemos verla a continuación:

El origen de HTML fue un sistema de hipertexto para compartir documentos electrónicos en 1980. La primera propuesta oficial para convertir **HTML** en un estándar se realizó en 1993. Aunque ninguna de las dos propuestas de estándar que se hicieron (**HTML** y **HTML+**) consiguieron convertirse en estándar oficial.

- **HTML 2.0** fue la primera versión oficial de HTML el IETF publicó el estándar en septiembre de 1995.
- **HTML 3.2** se publicó el 14 de Enero de 1997 por el W3C. Incorpora los applets de Java y texto alrededor de las imágenes.
- **HTML 4.0** se publicó el 24 de Abril de 1998. Entre las novedades que presenta se encuentran las hojas de estilos CSS y la posibilidad de incluir pequeños programas en las páginas web.
- **HTML 4.01** Se publicó el 24 de diciembre de 1999. Es una actualización de la versión anterior. En ese momento el [W3C](#) detuvo la actividad de estandarización de HTML hasta marzo de 2007, momento en que se retoma debido a la fuerza de las empresas que forman el grupo [WHATWG] (https://es.wikipedia.org/wiki/Web_Hypertext_Application_Technology_Working_Group "Ver la definición de "WHATWG"") y a la publicación de los borradores de **HTML 5.0**, que será la siguiente versión de este lenguaje.
- **HTML 5** es la versión más avanzada y la que se considera estándar actualmente aunque ha ido evolucionando por diferentes especificaciones (HTML 5.1 y HTML 5.2). Es una versión viva en la que se sigue trabajando y su estándar actual (HTML 5.2) fue definido el 14 de diciembre de 2017. Puedes encontrar las especificaciones de la última versión de HTML en el siguiente [enlace](#).

Tras la publicación del estándar HTML 4.01 se detecta su incompatibilidad con herramientas basadas en XML. Para evitar estos problemas se crea lenguaje XHTML que combina la sintaxis de HTML 4.0 con la de XML.

- **XHTML 1.0** fue la primera versión, se publicó el 26 de Enero de 2000. Es una adaptación de HTML 4.01 al lenguaje XML, por lo que mantiene sus características, y añade algunas restricciones y elementos de XML.
- La versión **XHTML 1.1** fue el primer intento de modularizar XHTML. Fue declarada como reemplazada el 27 de marzo de 2018.
- La versión **XHTML 2.0**, es la versión más actual aunque fue publicada el 16 de diciembre de 2010.

2. Estructura de un documento HTML

La estructura de una página HTML ha de ser coherente con la que hemos visto en el tema anterior para cualquier documento XML. Por ello tendrá un prólogo y un ejemplar.

- **Prólogo:** Todo documento HTML ha de tener una declaración del tipo de documento donde se le indica al navegador el tipo de documento que se va a iniciar y la versión de HTML utilizada para la codificación del mismo y, además, le permite interpretarlo correctamente.

Para la versión HTML 4.0, hay tres prólogos distintos que definen tres tipos de documentos HTML:

- **HTML 4.0 Strict.** Es la DTD utilizada por defecto con HTML 4.0. En estos documentos no se permite el uso de los elementos declarados *deprecated* en otras versiones o Recomendaciones HTML. La declaración del tipo de documento correspondiente es:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
```
- **HTML 4.0 Transitional.** Permite el uso de todos los elementos que permite el **HTML 4.0 Strict**, además de los elementos *deprecated*. La declaración del tipo de documento correspondiente es:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
```
- **HTML 4.0 Frameset.** Es una variante de **HTML 4.0 Transitional** para documentos que usan **frames**. En estos documentos el elemento **body** hay que reemplazarlo por un elemento **frameset**. La declaración del tipo de documento correspondiente es:

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

En el caso de **HTML5** solo existe una declaración de tipo de documento:

```
<!DOCTYPE html>
```

Ejemplar: En un documento HTML está delimitado por las etiquetas **<html>** y **</html>**. El ejemplar puede, a su vez dividirse en dos partes:

- **La cabecera**, delimitada por las etiquetas **<head>** y **</head>**.
 - Contiene la información sobre el título de la página, el autor, palabras clave, etc. Dentro de esta sección es obligatorio definir el título del documento, para ello se usan las etiquetas **<title>** **</title>**. Esta información no se presentará en la ventana del navegador, salvo el título que aparecerá en la barra de título de la parte superior.
- **El cuerpo**, contiene la información que se va a presentar en la pantalla.
 - Está limitado por las etiquetas **<body>** y **</body>**, salvo en los documentos de tipo **HTML 4.0 Frameset** donde éstas se sustituyen por **<frameset>** y **</frameset>**.

3. Identificación de etiquetas y atributos HTML

Un documento **HTML** está formado por etiquetas y atributos.

Al igual que en **XML** las etiquetas pueden ser de apertura, , o de cierre, . Una de las diferencias con **XML** es que la cantidad de etiquetas de **HTML** está limitada a aquellas que están definidas por el lenguaje.

Aunque **HTML** define una gran cantidad de etiquetas, estas no son suficientes para crear páginas complejas ya que la definición completa de ciertos elementos, como las imágenes y los enlaces, requiere información adicional. Como no es posible crear una etiqueta por cada elemento diferente, se añade la información adicional a las etiquetas mediante los **atributos** dando lugar a los **elementos**.

Para cada uno de los atributos hay definido un *conjunto de valores* que se le puede asignar, si el valor de un atributo no es válido, el navegador le ignora.

Cada una de las etiquetas **HTML** define los atributos que puede utilizar, aunque algunos de ellos son comunes a muchas etiquetas.

4. Clasificación de los atributos comunes según su funcionalidad

Atributos básicos: Se pueden usar en casi todas las etiquetas HTML.

Atributo	Descripción
name = "texto"	Permite asignar el nombre "texto" a un objeto HTML
title = "texto"	Asigna un título a un elemento HTML, mejorando así la accesibilidad. Dicho título es mostrado por los navegadores cuando el usuario pasa el ratón por encima del elemento. Es especialmente útil con los elementos: a, link, img, object, abbr y acronym
id = "texto"	Permite identificar al elemento HTML sobre el que se aplica de forma única mediante el identificador "texto". Sólo es útil cuando se trabaja con CSS y con Javascript. No pueden empezar por números y sólo puede contener letras, números, guiones medios y/o guiones bajos.
style = "texto"	Permite aplicar al elemento HTML el estilo "texto" directamente.
class = "texto"	Permite aplicar al elemento HTML el estilo "texto" definido en las CSS. No pueden empezar por números y sólo puede contener letras, números, guiones medios y/o guiones bajos.

Atributos para internacionalización: Los utilizan las páginas que muestran sus contenidos en varios idiomas o aquellas que quieren indicar de forma explícita el idioma de sus contenidos

Atributo	Descripción																
dir	Indica la dirección del texto por lo que sólo puede tomar dos valores: ltr (left to right) de izquierda a derecha. Es el valor por defecto. rtl (right to left) de derecha a izquierda.																
lang = "codigo"	Especifica el idioma del elemento mediante un código predefinido. Los posibles valores de este atributo se encuentran en el documento RFC 1766 , algunos de los valores posibles son: <table><tr><td>Código</td><td>Idioma</td><td>Código</td><td>Idioma</td></tr><tr><td>en</td><td>Inglés (Gran Bretaña)</td><td>es</td><td>Español</td></tr><tr><td>en-US</td><td>Inglés americano</td><td>fr</td><td>Francés</td></tr><tr><td>ja</td><td>Japones</td><td>fr-CA</td><td>Francés de Canada</td></tr></table>	Código	Idioma	Código	Idioma	en	Inglés (Gran Bretaña)	es	Español	en-US	Inglés americano	fr	Francés	ja	Japones	fr-CA	Francés de Canada
Código	Idioma	Código	Idioma														
en	Inglés (Gran Bretaña)	es	Español														
en-US	Inglés americano	fr	Francés														
ja	Japones	fr-CA	Francés de Canada														
xml:lang = "codigo"	Especifica el idioma del elemento mediante un código definido según la recomendación RFC 1766 .																

En las páginas **XHTML**, el atributo **xml:lang** tiene más prioridad que **lang** y es obligatorio incluirlo siempre que se incluye el atributo **lang**.

- Atributos de eventos y atributos para los elementos que pueden obtener el foco:** Sólo se utilizan en las páginas web dinámicas creadas con JavaScript. Como no es nuestro objetivo no lo vamos a contemplar.