

1. Clasificación de los lenguajes de marcas y ejemplos

Tipo	Descripción	Ejemplo
Marcado de presentación	Indica el formato del texto, es decir cómo ha de presentarse el documento. La etiquetas de marcado suelen estar ocultas al usuario.	Microsoft word
Marcado de Procedimiento	También está enfocado hacia la presentación del texto, sin embargo, las marcas que formatean el texto son visibles para el usuario y permiten procesamiento (realizar un conjunto de acciones) según el tipo de etiqueta.	HTML
Marcado descriptivo o semántico	Utiliza las marcas o etiquetas para describir los fragmentos de texto.	XHTML

2. Evolución histórica HTML

HTML es el acrónimo de HyperText Markup Language, que traducido a nuestro idioma es Lenguaje de Marcación de Texto.

- Es una herramienta para que el ordenador conectado a Internet interprete como visualizar el documento.
- Es un sistema de etiquetas que indica al ordenador cuando hay que señalar una cursiva, separar un párrafo o definir el color del texto.
- el HTML le da las indicaciones mencionadas al programa cliente, el browser o navegador para que presente el documento en la pantalla de la manera adecuada.

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 HTMLTags(Etiquetas HTML) y todavía hoy puede ser consultado online a modo de reliquia informática.

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 (en esta época se definieron las etiquetas para imágenes, tablas y formularios) 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 22 de septiembre de ese mismo año, 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). La versión HTML 3.2 se publicó el 14 de Enero de 1997 y es la primera recomendación de HTML publicada por el W3C. Esta revisión incorpora los últimos avances de las páginas web desarrolladas hasta 1996, como applets de Java y texto que fluye alrededor de las imágenes.

HTML 4.0 se publicó el 24 de Abril de 1998 (siendo una versión corregida de la publicación original del 18 de Diciembre de 1997) 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 el 24 de 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 futuro estándar HTML 5, cuyo primer borrador oficial se publicó el 22 de 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.

De forma paralela a su actividad con HTML, W3C ha continuado con la estandarización de XHTML, una versión avanzada de HTML y basada en XML. La primera versión de XHTML se denomina XHTML 1.0 y se publicó el 26 de Enero de 2000 (y posteriormente se revisó el 1 de Agosto de 2002).

XHTML 1.0 es una adaptación de HTML 4.01 al lenguaje XML, por lo que mantiene casi todas sus etiquetas y características, pero añade algunas restricciones y elementos propios de XML. La versión XHTML 1.1 ya ha sido publicada en forma de borrador y pretende modularizar XHTML. También ha sido publicado el borrador de XHTML 2.0, que supondrá un cambio muy importante respecto de las anteriores versiones de XHTML.

3. Comparación HTML frente XHTML

HTML y XHTML son las dos variantes que existen en el lenguaje HTML 5. Las diferencias entre esas dos variantes son básicamente sintácticas: la sintaxis HTML está inspirada en la norma SGML (aunque no la cumple estrictamente), mientras que la sintaxis XHTML está basada en la recomendación XML (aunque tampoco la cumple estrictamente).

En general, la sintaxis XHTML es más "estricta", en el sentido de imponer restricciones en la forma de escribir etiquetas, atributos o valores, mientras que la sintaxis HTML es más "flexible".

Si un documento contiene errores de sintaxis (HTML o XHTML) se dice que es un documento inválido. Si las páginas web se sirven al navegador como application/xhtml+xml, el navegador debe rechazar las páginas inválidas, pero si se sirven al navegador como text/html, los navegadores intentan mostrar el documento aunque contenga errores. Aunque los navegadores a menudo consiguen mostrar documentos inválidos, se aconseja validar y corregir los

documentos para asegurar que los navegadores puedan mostrarlos correctamente.

4. Ejemplo de un fichero de texto sin formato

Los archivos de texto plano (plain text), son aquellos formados exclusivamente por texto (sólo caracteres), sin ningún formato; es decir, no requieren ser interpretados para leerse (aunque pueden ser procesados en algunos casos). También son llamados archivos de texto llano, simple o sin formato.

Los archivos de texto llano carecen de información destinada a generar formatos (negritas, subrayado, cursivas, tamaño, etc.)² y tipos de letra (por ejemplo, Arial, Times, Courier, etc.).² Esta simplicidad permite que una gran variedad de programas pueda leer y editar ese contenido. Las aplicaciones destinadas a la escritura y modificación de archivos de texto se llaman editores de texto. Cada texto tiene un peso diferente porque tiene una codificación distinta.

5. Pon un ejemplo de un fichero XML

Este código XML genera la siguiente estructura:

- Pagina inicial (página de nivel superior por defecto)
- Mi Escuela (con las páginas secundarias Noticias, Anuncios y Trimestre)
- Clase (con páginas secundarias para cada año (2005, 2004, 2003, 2002 y 2001). Cada una de estas páginas secundarias tiene tres páginas secundarias: Informática, Matemáticas e Inglés)
- Agenda de Teléfonos (con las páginas secundarias Empleados (que a su vez tiene las páginas secundarias Profesores, Maestros y Soporte) y Estudiantes (con las páginas secundarias Actuales y Ex-alumnos))
- Manuales (con las páginas secundarias Políticas y Procedimientos (que a su vez tiene las páginas secundarias Empleados y Estudiantes), Ayuda Financiera y Formularios)
- Biblioteca (con las páginas secundarias Novedades, Manuales de Referencia y Acceso a Internet)
- Objetos Perdidos

```
<taxonomy>
  <page title="Mi Escuela">
    <page title="Trimestre" />
  </page>
  <page title="Clase">
    <page title="2005">
      <page title="Informática" />
      <page title="Matemáticas" />
      <page title="Inglés" />
    </page>
    <page title="2004">
      <page title="Informática" />
      <page title="Matemáticas" />
      <page title="Inglés" />
    </page>
    <page title="2003">
      <page title="Informática" />
      <page title="Matemáticas" />
      <page title="Inglés" />
    </page>
    <page title="2002">
      <page title="Informática" />
      <page title="Matemáticas" />
      <page title="Inglés" />
    </page>
    <page title="2001">
      <page title="Informática" />
      <page title="Matemáticas" />
      <page title="Inglés" />
    </page>
  </page>
  <page title="Agenda de Teléfonos">
    <page title="Empleados">
```

```
<page title="Profesores" />
<page title="Maestros" />
<page title="Soporte" />
</page>
<page title="Estudiantes">
  <page title="Actuales" />
  <page title="Ex-alumnos" />
</page>
</page>
<page title="Manuales">
  <page title="Políticas y Procedimientos">
    <page title="Empleados" />
    <page title="Estudiantes" />
  </page>
  <page title="Ayuda Financiera" />
  <page title="Formularios" />
</page>
<page title="Biblioteca">
  <page title="Novedades" />
  <page title="Manuales de Referencia" />
  <page title="Acceso a Internet" />
</page>
<page title="Objetos Perdidos">
</page>
</taxonomy>
```

6. Estructura básica de un documento HTML

Todos los documentos Html tienen la estructura que se muestra a continuación, aunque la etiqueta <body> puede ser sustituida por <frameset> para un tipo de páginas que dividen la ventana del navegador en varios cuadros (frames).

```
<HTML>
  <HEAD>
    <TITLE>Título de la página</TITLE>
    ...
  </HEAD>

  <BODY>
    Aquí iría el contenido de la página
  </BODY>
</HTML>
```

7. Estructura básica de un documento XHTML

La estructura básica de cualquier página o documento XHTML válido incluye al menos las siguientes etiquetas:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" lang="es" xml:lang="es">
<head>
  <title>...</title>
  ...
</head>
<body>
  ...
</body>
</html>
```

El DOCTYPE también puede ser:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

8. Comparación de XML con SGML

HTML y XML están basados en el lenguaje SGML que es el estándar de los lenguajes de Marcado. HTML sigue la norma SGML pero no sus principios. XML sigue el estándar SGML. Como referencia XHTML es HTML conforme a XML.

9. Comparación de XML con HTML

XML y HTML son lenguajes muy diferentes. Ambos nacen de la misma inspiración llamada SGML por lo que su sintaxis es similar, aunque cada uno fue diseñado para cumplir distintas funciones:

El XML (eXtensive Markup Language) es un lenguaje que fue concebido para describir información. Su función principal es ayudarnos a organizar contenidos y eso hace que los documentos XML sean portables hacia diferentes tipos de aplicaciones.

El HTML (HyperText Markup Language) por otro lado ha sido concebido para mostrar información, determinar cómo actúa y qué hace. Su función radica en ayudarnos a darle formato a los diversos contenidos de una página.

Resumiendo, el XML sirve para describir información y el HTML sirve para darle formato y presentarla a través de un navegador. O sea que el XML no es ni será nunca un reemplazo del HTML sino un complemento que sirve para manejar la información separada del formato.

10. Enumera varios programas que se pueden utilizar como visor de ficheros XML.

- Notepad++
- Google Drive
- Open Office