

Unidad Didáctica 1

El lenguaje HTML

Contenido

1. Introducción
2. HTML

1. Introducción

A pesar de la rápida evolución de la Web, HTML ha sido siempre la base de la misma y su avance ha condicionado a los demás elementos del sistema. Esto explica el hecho de que los navegadores se suelen clasificar dependiendo de la versión de HTML que sean capaces de interpretar.

2. HTML

HTML (*HyperText Markup Language*) no es un lenguaje de programación “verdadero”. En realidad su utilización está orientada a la descripción de documentos y los componentes que forman parte de los mismos. No obstante, (como se verá en la próximas páginas), tiene comandos y definiciones que no se visualizan pero que alteran la apariencia final de un documento.

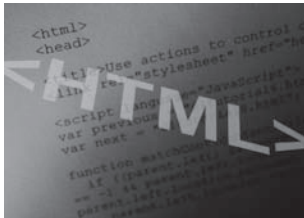
HTML es una versión simplificada del lenguaje SGML (*Standard Generalized Markup Language*), un estándar para la descripción de documentos. Es más, la gramática HTML está escrita en el lenguaje SGML (en un documento llamado DTO (*Document Type Definition*)).

SGML es un sistema de descripción de documentos, donde cada uno está organizado en una serie de entidades y lógicamente en una jerarquía de elementos. Provee una codificación estándar para la transmisión de documentos entre sistemas diferentes, como por ejemplo: distintas plataformas.

Un documento SGML se marca de tal modo que no define ningún aspecto relacionado con la representación (estilo). Para definir el aspecto, un programa debe de unir el documento SGML (marcado), con un esquema de estilo, a fin de producir una representación.

2.1. Las versiones de HTML

La mayoría de los lenguajes de programación permanecen, por lo general, estables en el tiempo y, en caso de sufrir alteraciones, cada versión queda perfectamente identificada respecto a las demás. La elaboración de una nueva



actualización es un proceso delicado que exige un acuerdo de los usuarios, el cual queda reflejado en las organizaciones de estandarización (ISO, ANSI, IEEE, ...).

Respecto al lenguaje HTML existen una serie de niveles definidos por el W3C (*World Wide Web Consortium*) que se corresponden con las diferentes versiones establecidas por este organismo:

El nivel 0 estableció la estructura básica de un documento con la etiqueta o contenedor principal `<HTML>` y las dos secciones principales: `<HEAD>` y `<BODY>`. En este nivel quedó definido el contenido de la primera y gran parte del de la segunda: estilos lógicos, listas, hipervínculos, imágenes, cabeceras, y el conjunto de caracteres.



Definición

Los estilos consisten en los distintos efectos que se pueden aplicar a un texto (negrita, cursiva, subrayado, etc.). En HTML existen dos tipos o grupos de estilos: los lógicos y los físicos. Los físicos indican efectos tales como negrita, cursiva, etc., y los lógicos indican un tipo de texto (cita, nombre de persona, etc.).

El nivel 1 añadió algunos estilos físicos y lógicos. Esta primera definición del lenguaje resultó insuficiente y se hizo necesario añadir nuevos comandos que permitieran mejorar la apariencia los documentos Web.

Cuando se construye un navegador se hace teniendo en cuenta la versión del lenguaje más reciente.

Si se intentara trabajar con versiones futuras más avanzadas de HTML sobre los visualizadores anteriores, no se produciría ningún error en el navegador, únicamente ignorará comandos no reconocidos. Por ejemplo, un navegador desarrollado para versiones anteriores, ignorará las etiquetas y comandos más recientes del lengua-

je (al ejecutar una página realizada para versiones posteriores) ya que no las reconocerá.

A la versión 1.0 siguió la 2.0, la cual implementaba el nivel 2 del lenguaje, cuya principal novedad podía localizarse en la incorporación de formularios.

El nivel 3, encarnado por HTML 3.0, se preparó como una evolución lógica del estándar. Sin embargo, la aparición del primer navegador de Netscape trajo consigo un sinnúmero de mejoras especiales que hizo necesaria la definición de un nuevo estándar, HTML 3.2 que se encargó de “legalizar” gran parte de estas mejoras, uniéndolas a la formalidad del modelo propuesto por el W3C.

La culminación de este proceso fue durante mucho tiempo HTML 4.0, propuesto por primera vez en 1997 y que trató de normalizar el lenguaje e incorporar ciertas mejoras propuestas por los fabricantes de navegadores. El estándar fue modificado ligeramente en 1999 y adoptó el nombre HTML 4.01 (o HTML 4.1). Éste fue, casi durante diez años, el último estándar de HTML.

2.2. XHTML

Antes de hablar de la última evolución de HTML (5), es necesario conocer el significado de XHTML.

XML (*eXtensible Markup Language*) es la culminación de los esfuerzos del W3C por sistematizar y regularizar la construcción de documentos. XML es un lenguaje de descripción de documentos que no incluye ninguna información relativa al diseño de éstos.

La idea consiste en dividir la información en dos partes: contenido (la información en bruto) y presentación. XML ha tenido un enorme éxito y ha sustituido a muchos estándares de descripción de información

Una de las características fundamentales de XML es que hace un control muy riguroso de la sintaxis de sus documentos, por lo que la tolerancia mencionada antes de los navegadores HTML es imposible con XML.

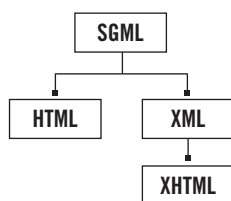
Tras cerrar el desarrollo de HTML, el W3C definió un estándar más estricto, basado en XML, que recibió el nombre de XHTML (de *eXtensible Hypertext Markup Language*). Durante mucho tiempo, XHTML se presentó como la evolución de HTML, el cual debería dejar de usarse en el futuro.

Sin embargo, los desarrolladores Web en general (páginas, navegadores etc.) descubrieron que suponía un gran esfuerzo y dificultad adaptar lo especificado en XHTML, por lo que apenas lo usaban.

Con el tiempo, el W3C aceptó que XHTML se convirtiera en una recomendación paralela que podía coexistir con HTML.

Aunque uno de los propósitos para definir XHTML consistía en facilitar la visualización de páginas en dispositivos diferentes al PC (móviles, PDAs). Finalmente, los navegadores de estos dispositivos han acabado soportando indistintamente HTML y XHTML.

Aunque la filosofía de separar contenido y presentación (principal aportación de XHTML) ha sido ampliamente aceptada, XHTML es un camino muerto cuya evolución se ha abandonado.



2.3. HTML 5

Tras varios años sin actividad, se constituyó (2004) el WHATWG (*Web Hypertext Application Technology Working Group*), al margen del W3C, y promovido por empresas como Apple, Opera, Google o la fundación Mozilla, y cuyo propósito ha sido la creación de una nueva versión del estándar desde un punto de vista fundamentalmente práctico, y no académico (como hasta ahora).

En 2007 el W3C reconoció este trabajo y lo tomó como base para su propia actividad. A finales de 2009 el WHATWG publicó la última versión.

Con esta nueva versión, HTML no cambia en lo esencial, aunque se enriquece con medios para simplificar el trabajo con las nuevas herramientas de gestión de contenidos (blogs, agregadores, páginas personales y de inicio, ...), y facilitar la inclusión de elementos multimedia.

El principal criterio de diseño de HTML 5 ha sido el de resolver problemas prácticos, lo que explica el esfuerzo de orientar esta versión a facilitar el trabajo en situaciones reales.

Uno de los principales cambios es la inclusión en el estándar del DOM, un componente fundamental de HTML, aunque siempre había sido tratado de forma separada.

Los navegadores disponibles actualmente (en el momento de redactar este manual) no soportan aún todos los nuevos elementos definidos en HTML 5. No obstante, el soporte a HTML 5 está aumentando muy rápidamente.