

# HTML (HyperText Markup Language)

## • História / Resumo

A linguagem HTML surgiu em 1993, a partir da combinação de duas linguagens: a SGML (Standard Generalized Markup Language) que suportava apenas elementos de texto e a HyTime (Hypermedia/Time-based Structuring Language) que, além de informações textuais, permitia o uso de conteúdo hipermídia (baseado em áudio e vídeo).

O consórcio W3C (World Wide Web Consortium) foi fundado em 1994 com o objetivo de padronizar a linguagem HTML. Até hoje o W3C especifica, controla e padroniza as linguagens web que os navegadores (browsers) conseguem interpretar e exibir. No seu site é possível testar páginas, validando a partir da URL do site, de Upload de Arquivos (File Upload) ou Cópia Direta (Direct Input). Site: <http://validator.w3.org>

Em 1997, o consórcio W3C apresenta oficialmente a linguagem HTML 4.0 com a possibilidade de usar folhas de estilo em cascata (CSS – Cascading Style Sheets) como linguagem de formatação de estilos e layouts.

## • Definições

Assim como qualquer linguagem utilizada para programar ações de um computador, a linguagem HTML é formada por um conjunto de palavras reservadas (conhecidas como "tags") que possibilitam estabelecer o trabalho que um computador vai realizar.

Um documento (página) HTML nada mais é que um arquivo de texto simples contendo códigos (comandos) baseados em etiquetas de identificação, também conhecidas pelo termo em inglês tags. Uma tag é uma palavra reservada delimitada pelos símbolos "<" (menor que) e ">" (maior que). Desta forma, as tags são os comandos de marcação da linguagem que especificam os elementos que estruturam uma página, como cabeçalhos, parágrafos, tabelas etc.

A programação de um documento baseado na linguagem HTML baseia-se somente em códigos (comandos especiais) de marcação de texto (tags), não tendo nenhuma complicação lógica, já que o documento é composto somente por linhas de texto e as tags dispostas de forma sequencial, respeitando as regras sintáticas da linguagem. Devido a essa característica, a construção de páginas com a linguagem HTML é considerada estática.

Os comandos HTML são interpretados pelo programa de navegação (browser). Quando o browser acessa uma página, ele procura os comandos da linguagem (tags) e os interpreta. Ao encontrar uma tag, o navegador verifica se é uma palavra reservada válida como comando da linguagem. Sendo válida, o navegador executa a ação indicada e apresenta o resultado. Se não for válida, o navegador não executa a ação e não apresenta nenhuma mensagem de erro.

As tags são geralmente usadas em duplas (tag de inicialização <tag> e tag de finalização </tag>), envolvendo o objeto (texto, imagem, link etc.) que será tratado. Existem algumas tags que são consideradas órfãs por não possuírem uma outra tag que indique finalização, devendo ser escritas como <tag />.

As tags podem ser grafadas em caracteres minúsculos ou maiúsculos, inclusive misturando-os, o que ocasiona um desagradável efeito visual e é considerado um péssimo hábito de programação. O consórcio W3C recomenda que as tags em HTML sejam escritas em caracteres **minúsculos**.

## • Estrutura Básica

A linguagem HTML possui uma estrutura básica (mínima) para a formatação de um documento (página). Ele deve conter algumas tags básicas para que o navegador entenda que a estrutura em uso deve ser interpretada por ele. Uma página HTML deve ser formada pelas estruturas: principal, cabeçalho, título do navegador e corpo do documento. Estas estruturas são identificadas pelas tags:

- <html>...</html> - Define a estrutura principal do documento. Marcam o início e o fim do documento. São obrigatórias, pois informam ao navegador que o documento está escrito na linguagem HTML. Possui outras duas partes distintas, sendo o cabeçalho (head) e o corpo (body) do documento.
- <head>...</head> - Permite definir o cabeçalho do documento. As informações contidas nestas tags não serão exibidas no corpo do documento. É dentro delas que se colocam o título que aparecerá na barra do navegador, os metadados para o reconhecimento do documento pelos programas de busca e pesquisa da Internet, os scripts que serão executados como funções do documento e o acesso a arquivos CSS externos.

- **<title>...</title>** - Permite definir o texto que será apresentado na barra de título do navegador (browser).
- **<body>...</body>** - Define o bloco no qual um documento HTML será estruturado.

A declaração **<!DOCTYPE>** não deve ser confundida com uma tag. Ela é expressa em caracteres maiúsculos e tem a finalidade de indicar ao navegador a linguagem e a versão usada para escrever o código do documento (página). É uma declaração obrigatória e deve sempre ser definida antes da tag **<html>**. Quando se utiliza a versão mais recente da linguagem HTML (**versão 5**), usa-se a declaração **<!DOCTYPE html>**. Mas, se estiver utilizando uma versão anterior, deve-se usar a forma completa que apresenta também o nome do formato da linguagem, sua versão, a especificação do tipo de documento e o endereço do local onde se encontra o arquivo de catálogo com a definição pública do código da página em uso.

Exemplo da declaração para HTML **versão 4.01**:

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

O uso das tags **<!--** e **-->** é opcional e permite inserir no código uma série de comentários explicativos para facilitar o entendimento do código. Este tipo de tag não é executada pelo navegador.

### • Criação das Páginas

A gravação do documento em HTML no programa de edição de textos deve usar uma extensão de identificação para reconhecimento do navegador. Neste caso, é ideal que essa extensão seja grafada como **.html**. No entanto, ainda é possível encontrar o uso da extensão **.htm** (que deve ser evitada), pois no passado os navegadores só reconheciam extensões de até três caracteres.

Aconselha-se que o formato de escrita obedeça ao uso e à definição de indentação, que é o deslocamento de espaços em branco dentro do código com o objetivo de tornar a leitura do código mais clara. A indentação é um artifício que possibilita uma estrutura do código mais leve e elegante aos olhos do desenvolvedor, facilitando a leitura do código e a implementação de correções ou alterações.

De forma geral, pode-se dizer que um site é um conjunto de páginas que estão relacionadas entre si com o objetivo de apresentar um conjunto de informações. É um excelente hábito procurar dividir sua estrutura em diretórios (pastas) para melhor organizar e distribuir seus arquivos. Por exemplo, as imagens podem estar armazenadas em um diretório que tenha um nome sugestivo. Também podem ser agrupados em diretórios com nomes sugestivos os documentos de texto, os arquivos de estilos e scripts.

Sugere-se o uso do nome **Index.html** para a página principal de um site, ou seja, a página que deve ser aberta primeiro quando o site for acessado.

Quando se cria um arquivo com extensão **.html** alguns editores de código podem fornecer a estrutura inicial da página. Para isso é necessário digitar um símbolo de exclamação **!** e pressionar ENTER ou TAB. Outras opções disponíveis, que variam de acordo com o editor, são as palavras **html** ou **html:5**

### • Meta Tags

As meta tags fornecem informações sobre a página ao navegador e aos serviços de busca na Internet. As meta tags são sempre declaradas entre as tags **<head>** e **</head>**.

Existe uma meta tag obrigatória que informa ao navegador o conjunto de caracteres padrão. Para o HTML 5, utilizamos o UTF-8. Ele inclui quase todos os caracteres e símbolos do mundo.

No HTML 4, esta meta tag era declarada como:

```
<meta http-equiv="content-type" content="text/html; charset=UTF-8">.
```

No HTML 5, utiliza-se a forma reduzida:

```
<meta charset="UTF-8">
```

Fonte:

MANZANO, José Augusto Navarro Garcia; TOLEDO, Suely Alves de. Guia de Orientação e Desenvolvimento de Sites HTML, XHTML, CSS e JavaScript/JScript, 2ª ed. São Paulo: Érica, 2010.