



HTML Overview

HTML (Hypertext Markup Language) is the markup language used to turn text documents into web pages. HTML allows authors to identify elements that give the document structure, such as headings, paragraphs, lists, and so on. Other elements act as mechanisms for adding hypertext links, interactive forms, and media such as audio and video to web pages.

HTML has undergone quite a journey since its creation by Tim Berners-Lee in 1991 as a simple way to indicate the meaning and structure of hypertext documents. It didn't take long for competing browser developers to add on to the initial minimal set of HTML elements or for the first crop of web designers to co-opt HTML as a visual design tool.

XHTML is a reformulation of HTML in XML. In other words, it uses the same vocabulary (the same elements and attributes) as HTML, but the syntactical rules are pulled from XML, which is stricter than HTML. XHTML is discussed in detail later in this chapter.

Before we delve into HTML and XHTML syntax, let's take a moment to look at the important role (X)HTML plays as well as the recent groundswell of respect it has earned in the new standards-driven web design environment.

The W3C

Seeing the need to bring order to the development of HTML, Berners-Lee founded the World Wide Web Consortium (W3C) in 1994. The W3C continues to oversee HTML and related web technologies and has been releasing updated and standardized versions of HTML in publications known as "Recommendations" since 1995. The current standards are HTML 4.01 (1999) and XHTML 1.0 (2000).

The Role of HTML

The marked up HTML document is said to be the *structural layer* of a web page. It is the foundation upon which the *presentation layer* (instructions for how the elements should be delivered or displayed) and the *behavioral layer* (scripting and interactivity) are applied.

Did you happen to read the preceding XML chapter? It may seem off the topic of HTML, but there are some critical XML-based concepts there that guide the way HTML is perceived and handled in contemporary web design. One guiding concept is that the fundamental purpose of HTML as a markup language is to provide a *semantic* description (the meaning) of the content and establish a document *structure*. It is not concerned with *presentation*, such as how the document will look in a browser. Presentation is the job of Cascading Style Sheets, which is covered in Part III.

That presentational instructions should be kept separate from the semantic and structural markup is nothing new. It has been the intent of HTML from its beginning as an application of SGML (Standardized General Markup Language) as noted in the upcoming sidebar. What is new is that the web community is recognizing that there are measurable advantages (in terms of time and money) to using HTML for what it was designed to do, and nothing more.

Keeping Presentation Separate from Document Structure

Before HTML, there was SGML (Standard Generalized Markup Language), which established a complex language for describing documents in terms of their structure, independent of appearance. SGML is a vast set of rules for developing markup languages such as HTML, but it is so all-encompassing that HTML uses only a small subset of its capabilities.

Because HTML is one instance of an SGML markup system, this principle of keeping presentation information separate from the structure of the document remains inherent to the HTML purpose. Over the early years of the Web's development, this ideal was compromised by the creation of HTML tags that contain explicit style instructions, such as the `font` element and `bgcolor` attribute.

The W3C has been taking measures to get HTML back on track. First, the creation of Cascading Style Sheets gives authors a robust solution for specifying style information and keeping it out of the document's content. In addition, with each new HTML Recommendation, the elements and attributes related to presentation have been deprecated and finally eliminated.

With this system in place, the W3C is more diligent than ever to clean up the HTML standard to make it work the way it was intended. Slowly the browser and authoring tool developers are getting on board. Now it is up to web developers and designers to start creating clean content.