

Was heißt „HTML“?

HTML bedeutet **HyperText Markup Language** und ist eine so genannte Auszeichnungssprache (Markup Language).

Eine Auszeichnungssprache (Markup Language) hat die Aufgabe, die logischen Bestandteile eines textorientierten Dokuments zu beschreiben. Als Auszeichnungssprache bietet HTML daher die Möglichkeit an, typische Elemente eines textorientierten Dokuments, wie Überschriften, Textabsätze, Listen, Tabellen oder Grafikreferenzen, als solche auszuzeichnen.

Das Auszeichnungsschema von HTML geht von einer hierarchischen Gliederung aus. HTML zeichnet Inhalte von Dokumenten aus. Dokumente haben globale Eigenschaften wie zum Beispiel Kopfdaten. Der eigentliche Inhalt besteht aus Elementen, zum Beispiel einer Überschrift 1. Ordnung, Textabsätzen, Tabellen und Grafiken. Einige dieser Elemente haben wiederum Unterelemente. So enthält ein Textabsatz zum Beispiel eine als betont oder fett markierte Textstelle, eine Aufzählungsliste besteht aus einzelnen Listenelementen, und eine Tabelle gliedert sich in einzelne Tabellenzellen.

HTML ist also eine Sprache zur Strukturierung von Texten, wobei wie erwähnt die Möglichkeit besteht, Grafiken und multimediale Inhalte in Form einer Referenz einzubinden und in den Text zu integrieren.

Die meisten dieser Elemente haben einen fest definierbaren Erstreckungsraum. So geht eine Überschrift vom ersten bis zum letzten Zeichen, eine Aufzählungsliste vom ersten bis zum letzten Listenelement, oder eine Tabelle von der ersten bis zur letzten Zelle. Auszeichnungen markieren Anfang und Ende von Elementen. Um etwa eine Überschrift auszuzeichnen, lautet das Schema:

```
<Überschrift> Text der Überschrift <Ende Überschrift>
```

Bei einem Element, das wiederum Unterelemente besitzt, etwa einer Aufzählungsliste, lässt sich das gleiche Schema anwenden:

```
<Liste>  
<Listenelement> Text des Listenelements <Ende Listenelement>  
<Listenelement> Text des Listenelements <Ende Listenelement>  
<Ende Liste>
```

Web-Browser, die HTML-Dateien am Bildschirm anzeigen, lösen die Auszeichnungsmarkierungen auf und stellen die Elemente dann in optisch gut erkennbarer Form am Bildschirm dar.

HTML allein kann weder Grafik-Designer befriedigen, die jeden Pixel am Bildschirm kontrollieren wollen, noch Daten-Designer, die aus der Welt der relationalen Datenbanken kommen und sich anwendungsspezifische Lösungen wünschen. Deshalb gibt es heute Style-Sprachen wie **CSS**, und es gibt Lösungen wie **XML**, um anwendungsspezifisches Daten-Design zu ermöglichen.

Mit Hilfe von Style-Sheets (CSS-Verfahren) lassen sich die unterschiedlichen Auszeichnungen in Ihrer Darstellung und Formatierung zentral definieren und auch Seiten-übergreifend einheitlich präsentieren.

Dabei ist die Bildschirmdarstellung aber nicht die einzige denkbare Ausgabeform. HTML kann beispielsweise genauso gut mit Hilfe synthetisch erzeugter Stimmen auf Audio-Systemen ausgegeben werden.

Das **W3-Konsortium** ist für die Standardisierung von HTML zuständig und bemüht sich, HTML von allen Sünden der Anfangsjahre zu reinigen und als einfache, reine Text-Strukturierungssprache zu etablieren.

Es ist durchaus möglich, HTML-4- oder XHTML-1-gerechte Web-Seiten zu erstellen, die richtig gut aussehen. Es gehört einfach ein wenig Kenntnis der Sachlage dazu.

HTML als software-unabhängiges Klartextformat

HTML ist ein so genanntes Klartextformat. HTML-Dateien können Sie mit jedem beliebigen Texteditor bearbeiten, der Daten als reine Textdateien abspeichern kann. Es gibt also keine bestimmte Software, die man zum Erstellen von HTML-Dateien benötigt. Zwar gibt es längst mächtige Programme, die auf das Editieren von HTML spezialisiert sind, doch das ändert nichts an der entscheidenden Eigenschaft: HTML ist nicht an irgendein bestimmtes, kommerzielles Software-Produkt gebunden.

Da HTML ein Klartextformat ist, lässt es sich auch hervorragend mit Hilfe von Programmen generieren. Von dieser Möglichkeit machen beispielsweise CGI-Skripts oder PHP Gebrauch. Wenn Sie im Web zum Beispiel einen Suchdienst benutzen und nach einer Suchanfrage die Ergebnisse präsentiert bekommen, dann ist das, was Sie am Bildschirm sehen, HTML-Code, der von einem Skript generiert wurde.

HTML für Hypertext

Eine der wichtigsten Eigenschaften von HTML ist die Möglichkeit, Verweise zu definieren. Verweise ("Hyperlinks") können zu anderen Stellen im eigenen Projekt führen, aber auch zu beliebigen anderen Adressen im World Wide Web und sogar zu Internet-Adressen, die nicht Teil des Web sind. Auf dieser Grundidee beruht letztlich das gesamte World Wide Web, und dieser Idee verdankt es seinen Namen.

Durch die „Vernetzung“ von Informationen mittels Hyperlinks entstehen so genannte „Hypertexte“. Hypertexte sind – im Gegensatz zu Büchern beispielsweise – nicht zum sequentiellen Lesen gedacht. Sie bieten also vielmehr Texteinheiten, die in sich abgeschlossen sind und jeweils einen kleinen "Informationshappen" anbieten. Durch die intelligente Vernetzung solcher Texteinheiten entsteht dann jener Informationsraum, den man als Hypertext bezeichnet.

HTML-Versionen

HTML ist eine Sprache, die mit Hilfe von SGML (Standard Generalized Markup Language) definiert wird. SGML gilt als „Mutter“ aller Auszeichnungssprachen und ist als ISO-Norm 8879 festgeschrieben. Mittlerweile gibt es einen Ableger von HTML namens XHTML. In der Version 1.0 ist XHTML eine Redefinition von HTML mit Hilfe von XML (Extensible Markup Language: Erweiterbare Auszeichnungssprache). XML hat als eine Untermenge von SGML nämlich ähnliche Aufgaben wie SGML selbst.

HTML 1.0 enthielt Auszeichnungen für Standard-Elemente wie Überschriften, Textabsätze, für Grafikreferenzen und natürlich für Verweise. Die erste Sprachversion von HTML gilt heute nicht mehr als erwähnenswert. Die Spezifikation ist auf den Seiten des W3-Konsortiums auch gar nicht mehr verfügbar.

HTML 2.0 wurde im November 1995 offizieller Sprachstandard und gilt heute als der allerkleinste gemeinsame Nenner. Allerdings enthält HTML 2.0 durchaus Sprachelemente, die bis auf den heutigen Tag von den großen Browsern nicht oder nicht vollständig unterstützt werden, etwa das `<link>`-Element. HTML 2.0 wurde allgemein als Enttäuschung empfunden, da gerade Netscape in seiner Entwicklung schon viel weiter war: So war HTML 2.0 Standard, als der Navigator in seiner überwältigenden Produktversion 2.0 auf den Markt kam. Darin waren bereits solche Dinge wie Frames (Mehrfenster-technik) und Multimedia-Referenzen implementiert, die meilenweit über den offiziellen HTML-Standard hinausgingen. Die meisten Ersteller von Web-Seiten berauschten sich damals an den neuen Möglichkeiten von Netscape und hatten wenig Verständnis für die Restriktionen von HTML 2.0.

HTML 3.2 wurde am 14.1.1997 offizieller Sprachstandard. Dass diese Version eine schwere Geburt war, lässt sich schon an der merkwürdigen Versionsnummer erahnen. Das W3-Konsortium, von den Erfolgen des Netscape-Browsers unbeeindruckt, entwickelte zunächst eine geplante Version 3.0. Die Vorschläge zu HTML 3.0 liefen völlig an den neuen Realitäten im Web vorbei und wurden nie offiziell. HTML 3.2 ist das Ergebnis einer völligen Umarbeitung. Endlich wurden zumindest Tabellen offizieller Bestandteil von HTML. Daneben aber auch einige Elemente zur physischen Textauszeichnung. Deshalb stellt HTML 3.2 aus heutiger Sicht ein mittleres Unglück für die Entwicklung von HTML dar. Denn in dieser Version wurde versucht, HTML als Sprache für optisches Design zu etablieren, was jedoch nicht dem Wesen der Sprache entsprach und auch nur halbherzig implementiert war. Viele Bestandteile, die in HTML 3.2 eingeführt wurden, sind mittlerweile als *deprecated* (missbilligt) eingestuft und sollen künftig wieder aus dem HTML-Sprachstandard entfallen, weil sie durch andere Techniken wie CSS realisierbar sind.

HTML 4.0 wurde erstmals am 18.2.1998 als Sprachstandard verabschiedet. Diese Sprachversion wurde jedoch mehrfach überarbeitet und liegt mittlerweile in einer Version 4.01 vor. HTML 4.0 stellt den Versuch dar, sich auf die Kernaufgaben von HTML zurückzubesinnen, und regelt vor allem die Einbindung von CSS Style Sheets und von Script-Sprachen wie JavaScript in HTML. Ein weiterer Schwerpunkt ist die Internationalisierung. HTML 4.0 setzt im Gegensatz zu seinen Vorläufern voll auf das Unicode-System und erlaubt damit, Texte in allen nur erdenklichen Sprachen der Welt HTML-gerecht zu notieren. Es gibt allerdings auch viel Verwirrendes in HTML 4.0. Dazu gehören vor allem die drei Sprachvarianten, die das W3-Konsortium anbietet: "HTML Strict", "HTML Transitional" und "HTML Frameset". Für Viele ist das Konzept, das hinter dieser Aufteilung steckt, kaum nachvollziehbar.

Kompliziert wurde die Sache auch dadurch, dass HTML - im Zuge der Etablierung von XML - in Gestalt von XHTML neu definiert wurde. So existiert heute neben dem SGML-basierten, klassischen HTML das XML-basierte XHTML. Letzteres ist zwar auch "nur" HTML, aber aufgrund XML-bedingter Gegebenheiten gibt es einige Unterschiede zu HTML. Von XHTML gibt es mittlerweile auch schon zwei Versionen: eine Version 1.0 und eine Version 1.1.