

Hypertext Markup Language

aus Wikipedia, der freien Enzyklopädie

Die **Hypertext Markup Language** (engl. ‚Hypertext-Auszeichnungssprache‘), abgekürzt **HTML**, ist eine textbasierte Auszeichnungssprache zur Strukturierung von Inhalten wie Texten, Bildern und Hyperlinks in Dokumenten. HTML-Dokumente sind die Grundlage des World Wide Web und werden von einem Webbrowser dargestellt. Neben den vom Browser angezeigten Inhalten einer Webseite enthält HTML zusätzliche Angaben in Form von Metainformationen, die z. B. über die im Text verwendete Sprache oder den Autor Auskunft geben oder den Inhalt des Textes zusammenfassen.

Die Auszeichnungssprache wird vom World Wide Web Consortium (W3C) und der Web Hypertext Application Technology Working Group (WHATWG) weiterentwickelt. Aktuell trägt HTML die Versionsnummer 4.01. HTML5 des W3C bzw. der „HTML Living Standard“ der WHATWG befindet sich in der Entwicklung, wird jedoch schon von vielen aktuellen Layout-Engines unterstützt. Parallel existiert die Extensible Hypertext Markup Language (XHTML).

HTML (Hypertext Markup Language)

```
<!DOCTYPE html>
<html>
<!-- created 2010-01-01 -->
<head>
<title>sample</title>
</head>
<body>
<p>Voluptatem accusantium
totam rem aperiam.</p>
</body>
</html>
```

HTML

Dateiendung: .html, .htm

MIME-Type: text/html

Entwickelt von: World Wide Web Consortium

Aktuelle Version: 4.01

Art: Auszeichnungssprache

Erweitert zu: XHTML, HTML5

Standard(s): ISO/IEC 15445;

W3C HTML 4.01,^[1] W3C HTML 3.2^[2]

Website: www.w3.org/html
(<http://www.w3.org/html/>)

Inhaltsverzeichnis

- 1 Entstehung
- 2 Syntax
- 3 Sprachtyp
- 4 Versionen
- 5 HTML-Struktur
 - 5.1 Allgemeine Struktur
 - 5.2 HTML-Kopf
 - 5.3 HTML-Körper
- 6 HTML-Varianten
 - 6.1 Strict
 - 6.2 Transitional
 - 6.3 Frameset
- 7 Zusatztechniken und Weiterentwicklungen
 - 7.1 Cascading Style Sheets
 - 7.2 Dynamisches HTML

- 7.3 XML
- 7.4 Ajax
- 8 Unterschiedliche Darstellung und Interpretation von HTML-Dokumenten
- 9 Siehe auch
- 10 Literatur
- 11 Weblinks
 - 11.1 Tutorials
 - 11.2 Validierung
 - 11.3 Standards
- 12 Einzelnachweise

Entstehung

Vor der Entwicklung des World Wide Web und dessen Bestandteilen, zu denen auch HTML gehört, war es nicht möglich, Informationen auf digitalem Weg einfach, schnell und strukturiert zwischen mehreren Personen auszutauschen. Man benötigte neben Übertragungsprotokollen auch eine einfach zu verstehende Textauszeichnungssprache. Genau hier lag der Ansatzpunkt von HTML. Um Forschungsergebnisse mit anderen Mitarbeitern der Europäischen Organisation für Kernforschung (CERN) zu teilen und von den beiden Standorten in Frankreich und in der Schweiz aus zugänglich zu machen, entstand 1989 am CERN ein Projekt, welches sich mit der Lösung dieser Aufgabe beschäftigte. Am 3. November 1992 erschien die erste Version der HTML-Spezifikation.

Syntax

Dem Text wird durch Auszeichnungen (englisch *markup*) von Textteilen eine Struktur verliehen. Die Auszeichnung erfolgt durch SGML-Elemente. Die meisten dieser HTML-Elemente werden durch ein Tag-Paar markiert, das heißt durch einen Starttag und einen Endtag. Ein Starttag beginnt immer mit dem Zeichen <. Es folgt der Elementname (z. B. p für einen Absatz oder h1 für eine Überschrift erster Ordnung) und optional eine Liste seiner Attribute (z. B. class="warning"). Mit einem > wird der Starttag geschlossen. Ein Endtag besteht aus den Zeichen </, dem Elementnamen und dem abschließenden >. Die zusammengehörenden Start- und Endtags bilden zusammen mit dem dazwischenliegenden Inhalt ein Element allgemeiner SGML-Spezifikation. Diese Elemente lassen sich nach Regeln, die in einer Dokumenttypdefinition (DTD) angegeben sind, verschachteln:

```
<p>Ein Textabsatz, der ein <em>betontes</em> Wort enthält.</p>
```

Bestimmte Elemente müssen nicht explizit notiert werden. Bei einigen Elementen darf gemäß der SGML-Regel "OMITTAG"^[3] der Endtag fehlen (z. B. </p> oder). Zudem spielt bei Element- und Attributnamen Groß- und Kleinschreibung keine Rolle (z. B. , ,). Zum Vergleich: In XHTML sind diese Regeln strenger verfasst.

Neben Elementen mit Start- und Endtag gibt es in HTML auch inhaltsleere Elemente, wie etwa Zeilenumbrüche (br) oder Bilder (img).

```
Eine Textzeile,<br>die hier fortgesetzt wird.  

```

Es geht in HTML um beschreibende (englisch *descriptive*), nicht um verfahrens- (englisch *procedural*) und darstellungsorientierte (englisch *presentational*) Textauszeichnung, auch wenn sich HTML in früheren Versionen dafür verwenden ließ. HTML-Elemente sind keine Angaben zur Präsentation, die dem Webbrowser mitteilen, wie er den Text visuell zu formatieren hat. Vielmehr sind Elemente eine strukturierende Auszeichnung, mit der sich Textbereichen eine Bedeutung zuordnen lässt, z. B. `<h1>...</h1>` für eine Überschrift, `<p>...</p>` für einen Textabsatz und `...` für betonten Text. Wie diese Bedeutung letztlich dem Benutzer vermittelt wird (im Falle einer Überschrift z. B. durch vergrößerte, fette Schrift), ist zunächst dem Webbrowser überlassen und hängt von der Ausgabe-Umgebung ab. Denn obwohl HTML-Dokumente in der Regel auf Computerbildschirmen dargestellt werden, können sie auch auf anderen Medien ausgegeben werden, etwa auf Papier oder mittels Sprachausgabe. CSS-Formatvorlagen eignen sich dazu, um auf die Präsentation eines HTML-Dokuments in verschiedenen Medien Einfluss zu nehmen.

Daher gelten Elemente und Attribute zur Präsentation wie `...`, `<u>...</u>` und `noshade` als missbilligt (englisch *deprecated*) und sollten nach allgemeiner Auffassung vermieden werden.

Das Einlesen des Quelltextes sowie das Verarbeiten der vorhandenen Informationen wird in der Fachsprache auch als *Parsen* bezeichnet, und die Aufbereitung für das Ausgabemedium als *Rendern*. Die Sprache *HTML* beschreibt, wie der Browser (oder ein anderes Programm, wie z. B. ein Text-Editor) die Auszeichnungen des Textes zu „verstehen“ hat, nicht, wie er sie dann in der Darstellung umsetzt. So besagt `<h1>` zwar, dass eine Überschrift folgt, nicht aber, in welcher Schriftgröße oder Schriftschnitt diese darzustellen ist – hier haben sich nur gewisse übliche Standardeinstellungen eingebürgert, die aber nicht Teil der HTML-Spezifikation sind.

Sprachtyp

HTML ist eine Auszeichnungssprache und wird als solche auch nicht *programmiert*, sondern schlicht *geschrieben*.

Ein ähnliches Konzept (logische Beschreibung) wie hinter HTML steht hinter dem Satzsystem TeX/LaTeX, das im Unterschied zu HTML jedoch auf die Ausgabe per Drucker auf Papier zielt.

Versionen

HTML wurde am 13. März 1989 von Tim Berners-Lee am CERN in Genf festgelegt.^[4]

- **HTML** (ohne Versionsnummer, 3. November 1992): Die Urversion, die sich nur an Text orientierte.^[5]
- **HTML** (ohne Versionsnummer, 30. April 1993): Zu Text kommt neben Attributen wie fette oder kursive Darstellung die Bildintegration dazu.
- **HTML+** (November 1993) Geplante Erweiterungen, die in spätere Versionen einfließen, aber nie als HTML+ verabschiedet wurden.^[6]
- **HTML 2.0** (November 1995): Die mit RFC 1866 definierte Version führt u. a. Formulartechnik ein. Der Status dieses Standards ist „HISTORIC“. Auch die Vorgänger sind veraltet.
- **HTML 3.0**: Die Version erscheint nicht, weil sie mit der Einführung des Netscape-Browsers in der Version 3 bereits vor der Veröffentlichung veraltet ist.
- **HTML 3.2** (14. Januar 1997): Neu in dieser Version sind zahlreiche Features wie Tabellen, Textfluss um Bilder, Einbindung von Applets.
- **HTML 4.0** (18. Dezember 1997): Mit dieser Version werden Stylesheets, Skripte und Frames eingeführt. Auch eine Trennung in *Strict*, *Frameset* und *Transitional* erfolgt. Am 24. April 1998 erscheint eine leicht korrigierte Version dieses Standards.

- **HTML 4.01** (24. Dezember 1999): Ersetzt HTML 4.0 mit vielen kleineren Korrekturen.
- **XHTML 1.0** (26. Januar 2000): Eine Neuformulierung von HTML 4.01 mit Hilfe von XML. Am 1. August 2002 erscheint eine überarbeitete Version.
- **XHTML 1.1** (31. Mai 2001): Nachdem XHTML in Module aufgeteilt wurde, wird mit XHTML 1.1 eine strikte Version definiert, bei der die mit HTML 4 eingeführten Varianten Frameset und Transitional entfallen.
- **XHTML 2.0** (closed, ^[7] 26. Juli 2006): Diese Version sollte nicht mehr auf HTML 4.01 basieren und einige neue Elemente einführen, so z. B. `<n1>` für Navigationslisten. Die Trennung von Auszeichnung und Stil sollte in dieser Version vollendet werden. – Das W3C hat die Arbeiten an XHTML 2.0 im Sommer 2009 zu Gunsten von HTML5 eingestellt. ^{[8][9]}
- **HTML5** (Working Draft, ^[10] 23. April 2009): HTML5 schafft auf Basis von HTML 4.01 und XHTML 1.0 ein neues Vokabular. Die zu HTML gehörende DOM-Spezifikation wird ebenfalls überarbeitet und erweitert.

HTML-Struktur

Allgemeine Struktur

Ein HTML-Dokument besteht aus drei Bereichen:

1. der Dokumenttypdeklaration (Doctype) ganz am Anfang der Datei, die die verwendete Dokumenttypdefinition (DTD) angibt, z. B. HTML 4.01 Strict,
2. dem HTML-Kopf (HEAD), der hauptsächlich technische oder dokumentarische Informationen enthält, die üblicherweise nicht im Anzeigebereich des Browsers dargestellt werden
3. dem HTML-Körper (BODY), der jene Informationen enthält, die gewöhnlicherweise im Anzeigebereich des Browsers zu sehen sind.

Somit sieht die Grundstruktur einer Webseite wie folgt aus:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
<html>
  <head>
    <title>Titel der Webseite</title>
    <!-- weitere Kopfinformationen -->
    <!-- Kommentare werden im Browser nicht angezeigt. -->
  </head>
  <body>
    <p>Inhalt der Webseite</p>
  </body>
</html>
```

HTML-Kopf

Im Kopf können sieben verschiedene Elemente verwendet werden:

title

bezeichnet den Titel der Seite, der von den meisten Browsern in der Titelleiste angezeigt wird.

meta

kann vielfältige Metadaten enthalten. Siehe Meta-Element.

base

gibt entweder eine Basis-URI oder einen Basisframe an.

- link**
dient zur Angabe von logischen Beziehungen zu anderen Ressourcen. Wird am häufigsten zur Einbindung von Stylesheets benutzt.
- script**
bindet Code in einer bestimmten Skriptsprache ein, hauptsächlich JavaScript.
- style**
enthält Stilinformationen, hauptsächlich CSS-Deklarationen.
- object**
bindet eine externe Datei ein. Browser dürfen solche Objekte im Dokumentkopf nicht darstellen.

HTML-Körper

HTML unterscheidet zwischen Block- und Inline-Elementen. Der wesentliche Unterschied ist, dass erstere in der Ausgabe einen eigenen Block erzeugen, in dem der Inhalt untergebracht wird, während die Inline-Elemente den Textfluss nicht unterbrechen. Vereinfacht gesprochen haben Block-Elemente immer ihre eigene Zeile. Mithilfe von CSS ist es jedoch möglich, Block-Elemente wie ein Inline-Element darzustellen und umgekehrt. Zudem lassen sich alle Elemente via CSS auch als *inline-block* auszeichnen, mit dem Ergebnis, dass ein solches Element sowohl Eigenschaften eines Block-Elementes als auch eines Inline-Elementes besitzt.

Eine Überschrift erster Ordnung wird so ausgezeichnet:

```
<h1>Überschrift</h1>
```

h1 steht für *Heading 1*, zeichnet also eine Überschrift der ersten (und in HTML höchsten) Gliederungsstufe aus. Weiter möglich sind h2 bis h6, Überschriften zweiter bis sechster Gliederungsstufe.

Ein Hyperlink:

```
<a href="http://example.com/">Dies ist ein Verweis auf example.com</a>
```

Hyperlinks sind Verweise auf andere Ressourcen, meistens ebenfalls HTML-Dokumente, die üblicherweise im Browser durch Klick verfolgt werden können. Dieser Link könnte so gerendert werden: Dies ist ein Verweis auf example.com (<http://example.com/>) Ebenso ist an diesem Beispiel zu sehen, dass das Link-Element ein Inline-Element ist und keine neue Zeile beginnt.

Zur Logik stehen zum Beispiel die Elemente `strong` oder `em` bereit, mit denen sich stark hervorgehobener oder betonter Text auszeichnen lässt. Per Voreinstellung (lt. W3C-Empfehlung) werden `strong`- und `em`-Elemente durch **Fettschrift** beziehungsweise *kursive Schrift* gerendert.

Die Strukturbeschreibung des Textes, wie sie die obigen Beispiele veranschaulichen, vereinfacht es, das Rendern dem Betrachter anzupassen, um etwa Text einem Sehbehinderten vorzulesen oder als Braille auszugeben.

HTML-Varianten

Beim Entwurf der letzten HTML-Version 4 sollte der Tatsache, dass in vielen HTML-Dokumenten noch Elemente und Attribute zur Präsentation eingesetzt werden, Rechnung getragen werden. Das Ergebnis waren schließlich drei Varianten:

Strict

Diese DTD umfasst den Kernbestand an Elementen und Attributen. Es fehlen die meisten Elemente und Attribute zur Beeinflussung der Präsentation, unter anderem die Elemente *font*, *center* und *u* sowie Attribute wie *bgcolor*, *align* und *target*. Deren Rolle sollen in *Strict*-Dokumenten Stylesheets übernehmen. Text und nicht-blockbildende Elemente innerhalb der Elemente *body*, *form*, *blockquote* und *noscript* müssen sich grundsätzlich innerhalb eines Container-Elements befinden, zum Beispiel in einem *p*-Element.

Transitional

Die Transitional-Variante enthält noch ältere Elemente und Attribute, die auch *physische* Textauszeichnung ermöglichen. Durch diese DTD soll Webautoren, die noch nicht logische Strukturierung und Präsentation voneinander trennen, die Möglichkeit gegeben werden, standardkonformes HTML zu schreiben. Gleichzeitig soll sie sicherstellen, dass bestehende Webseiten weiterhin durch aktuelle Webbrowser angezeigt werden können.

Frameset

Diese Variante enthält zusätzlich zu allen Elementen der Transitional-Variante noch die Elemente für die Erzeugung von Framesets.

Zusatztechniken und Weiterentwicklungen

Cascading Style Sheets

Im Laufe der Jahre ist HTML um Elemente erweitert worden, die sich mit der Gestaltung des Dokuments befassen, was der ursprünglichen Idee der Systemunabhängigkeit entgegenlief. Eine Rückbesinnung auf die Trennung von Struktur und Layout (besser: Präsentation) wurde durch die Definition von Cascading Style Sheets (CSS) vorgenommen. So soll das Aussehen bzw. die Darstellung des Dokuments in einer separaten Datei, dem sogenannten Stylesheet, festgelegt werden. Dies verbessert die Anpassungsfähigkeit des Layouts an das jeweilige Ausgabegerät und an spezielle Bedürfnisse der Benutzer, beispielsweise eine spezielle Darstellung für Sehbehinderte. Heutzutage ist die CSS-Unterstützung der Browser ausreichend, um damit eine anspruchsvolle Gestaltung zu realisieren.

Dynamisches HTML

Schon sehr früh in der Geschichte von HTML wurden Zusatztechniken erfunden, die es ermöglichen, HTML-Dokumente während der Anzeige im Browser dynamisch zu verändern. Die gebräuchlichste ist JavaScript. Man spricht bei solchen interaktiven Dokumenten von dynamischem HTML. Diese Techniken wurden von verschiedenen Browser-Herstellern, allen voran Microsoft und Netscape, unabhängig voneinander entwickelt. Daher gab es erhebliche Probleme bei der Umsetzung der Techniken zwischen den verschiedenen Browsern. Mittlerweile interpretieren alle verbreiteten JavaScript-fähigen Browser das Document Object Model (DOM). Dadurch ist es möglich, in allen Browsern lauffähige Skripte zu schreiben. Es gibt jedoch noch immer Differenzen bei der Unterstützung des DOM-Standards.

XML

Version HTML 4.01 der HTML-Empfehlung wurde in der Metasprache XML neu formuliert. Das daraus entstandene XHTML 1.0 genügt den im Vergleich zu SGML strengeren syntaktischen Regeln von XML, ist aber in seinen drei DTD-Varianten semantisch mit der jeweils entsprechenden DTD-Variante von HTML 4.01 identisch. XHTML wird im Rahmen des HTML5 Standards weiter entwickelt.

Ajax

Mit der Ajax-Technologie ist es mittels JavaScript möglich, einzelne bereits geladene Webbrowser-Inhalte gezielt zu ändern und nachzuladen, ohne dass die Webseite komplett neu geladen werden muss. Wegen des geringeren Datenaufkommens wird zum einen eine schnellere Webserver-Antwort ermöglicht, und zum anderen lassen sich Reaktionsweisen von Desktop-Anwendungen simulieren.

Unterschiedliche Darstellung und Interpretation von HTML-Dokumenten

HTML dient als Auszeichnungssprache dazu, einen Text semantisch zu strukturieren, nicht aber zu formatieren. ^[1] Die Darstellung ist kein Teil der entsprechenden Spezifikationen und ist somit willkürlich vom Browseranbieter bestimmbar. Eine Ausnahme stellen die als deprecated markierten, präsentationsbezogenen Elemente dar.

Siehe auch

- Entitäten
- MIME Encapsulation of Aggregate HTML Documents (MHTML)
- Liste von HTML-Editoren

Literatur

- Stefan Münz, Wolfgang Nefzger: *HTML Handbuch*. Franzis-Verlag, Poing 2005, ISBN 3-7723-6654-6.
- Stefan Mintert (Hrsg.): *XHTML, CSS & Co. Die W3C-Spezifikationen für das Web-Publishing*. Addison-Wesley, München 2003, ISBN 3-8273-1872-6.
- Mark Lubkowitz: *Webseiten programmieren und gestalten – HTML, CSS, JavaScript, PHP, Perl, MySQL, SVG und Newsfeeds, mit CD*. Galileo Press, Bonn 2004, ISBN 3-89842-557-6.
- Elisabeth Robson & Eric Freeman: *HTML und CSS von Kopf bis Fuß*, O'Reilly, Köln 2012, ISBN 978-3-86899-934-1

Weblinks

 **Commons: HTML** ([//commons.wikimedia.org/wiki/Category:HTML?uselang=de](https://commons.wikimedia.org/wiki/Category:HTML?uselang=de)) – Sammlung von Bildern, Videos und Audiodateien

- Einige Notizen zum Ursprung von HTML (<http://www.w3.org/MarkUp/historical>) (englisch)
- Webpräsenz des Standardisierungsgremiums W3C (<http://www.w3.org/MarkUp/>) – Spezifikationen und Neuigkeiten (englisch)

Tutorials

 **Wikibooks: Websiteentwicklung** – Lern- und Lehrmaterialien

 **Wikibooks: Handbuch Webdesign** – Lern- und Lehrmaterialien

- SELFHTML (<http://de.selfhtml.org/>) – Ausführliches Tutorial und Dokumentation, die sich mit der Erstellung von Webseiten befasst.
- jendryschik.de: Einführung in XHTML, CSS und Webdesign (<http://jendryschik.de/wsdev/einfuehrung/>)

Validierung

- Checkliste für Webstandards (<http://www.lingo4u.de/article/checklist>)
- W3C-Validator (<http://validator.w3.org/>) zur Überprüfung der syntaktischen Richtigkeit eines HTML-Dokuments (englisch)
- Verschiedene nützliche Testwerkzeuge (<http://uitest.com/de/analysis/>), Site Check (<http://uitest.com/en/check/>) (englisch)
- WDG Validator (<http://www.htmlhelp.com/tools/validator/>) zur Überprüfung einer kompletten Webseite

Standards

- edition W3C.de (<http://www.edition-w3c.de/>) Die W3C-Spezifikationen in deutscher Übersetzung und Kommentierung
- HTML-Standards: 2.0 (<http://www.w3.org/TR/WD-html2/>), 3.2 (<http://www.w3.org/TR/REC-html32>), 4.0 (<http://www.w3.org/TR/1998/REC-html40-19980424/>), 4.01 (<http://www.w3.org/TR/html4/>) (englisch)

Einzelnachweise

1. <http://www.w3.org/TR/1999/REC-html401-19991224/> W3C HTML 4.01
2. <http://www.w3.org/TR/REC-html32-19970114> W3C HTML 3.2
3. SGML and HTML Explained; Chapter 9 Tag Minimization; 9.2 Tag omission (<http://www.is-thought.co.uk/book/sgml-9.htm#OMITTAG>) (englisch)
4. *Information Management: A Proposal*, Tim Berners-Lee, 1989 (<http://www.w3.org/History/1989/proposal.html>)
5. *HTML*, Urversion (<http://www.w3.org/History/19921103-hypertext/hypertext/WWW/MarkUp/MarkUp.html>)
6. *A Review of the HTML+ Document Format*, David Raggett (http://www.w3.org/MarkUp/htmlplus_paper/htmlplus.html)
7. *XHTML 2.0, W3C Working Draft* (<http://www.w3.org/TR/xhtml2/>)
8. Heise-Meldung vom 3. Juli 2009: *XHTML 2 wird eingestellt* (<http://www.heise.de/newsticker/XHTML-2-wird-eingestellt-/meldung/141514>)
9. Meldung des W3C zur Einstellung der Entwicklung von XHTML 2 (<http://www.w3.org/News/2009#item119>)
10. *HTML5, W3C Working Draft* (<http://www.w3.org/TR/html5/>)
11. HTML als Auszeichnungssprache (<http://de.selfhtml.org/intro/technologien/html.htm#auszeichnungssprache>)

Normdaten (Sachbegriff): GND: 4373477-7

Von „http://de.wikipedia.org/w/index.php?title=Hypertext_Markup_Language&oldid=120365368“

Kategorien: [Offenes Format](#) | [Web-Entwicklung](#) | [HTML](#) | [Beschreibungssprache](#)

- Diese Seite wurde zuletzt am 9. Juli 2013 um 05:34 Uhr geändert.
- Abrufstatistik

Der Text ist unter der Lizenz „Creative Commons Attribution/Share Alike“ verfügbar; Informationen zum Lizenzstatus eingebundener Mediendateien (etwa Bilder oder Videos) können im Regelfall durch Anklicken dieser abgerufen werden. Möglicherweise unterliegen die Inhalte jeweils zusätzlichen Bedingungen. Durch die Nutzung dieser Website erklären Sie sich mit den Nutzungsbedingungen und der Datenschutzrichtlinie einverstanden.

Wikipedia® ist eine eingetragene Marke der Wikimedia Foundation Inc.