

LaTeX und HTML

Teil 1

Günter Partosch, HRZ Gießen

E-Mail: `Guenter.Partosch@hrz.uni-giessen.de`

Inhalt des Vortrags

- HTML -- eine Dokumentenbeschreibungssprache
- HTML-Dokumente -- eine neue Dokumentenart
- Vergleichende Darstellung von LaTeX und HTML
 - ♦ Allgemeines
 - The Name of the Game
 - Eingabedatei
 - ♦ Setzen bzw. Ausgabe von Text
 - Zeilen- und Seitenumbruch
 - Spezielle Zeichen
 - Abstände
 - Kapitel und Überschriften
 - Fußnoten
 - hervorgehobene Textteile
 - Behandlung spezieller Textteile
 - Querverweise innerhalb des Dokuments
 - ♦ Spezialitäten
 - Schriftarten und Schriftgrößen
 - Abstände
 - Style-Sheets
 - ♦ „Übriggebliebenes“
 - In LaTeX: Formeln, kompliziertere Tabellen
 - In HTML: Hypertext-Strukturen, Formulare

HTML – eine Dokumentenbeschreibungssprache

- HTML (Hypertext Markup Language) ist SGML-Applikation
- übernimmt daher sehr viele Eigenschaften von HTML (z.B. logisches Markup, Entities)
- HTML wird durch eine DTD (Document Type Description) definiert
- Norm sagt fast nichts über die tatsächliche Darstellung von Textteilen aus.
- es gibt Validierer, die ein HTML-Dokument auf Korrektheit überprüfen
- sehr viele firmenspezifische Entwicklungen
- schnell voranschreitende Entwicklung geht einher mit abnehmender Akzeptanz der Normierung

HTML-Versionen:

- 2.0: (noch) die meist eingesetzte Version; unterstützt schon Bilder, interaktive Formulare, minimale typographische Elemente
- 2.1: minimale Erweiterungen
- 2.2: beinhaltet Version 2.1 und Tabellen aus Version 3
- 3.0: zusätzliche Zeichen- und Absatzformatierungen
- 3.1: Tabellen und mathematische Formeln
- 3.2: verwirft Version 3.0/3.1; zusätzlich Tabellen (einfacheres Modell), Applets, Hoch-/Tiefstellung, Textfluß um Bilder, Zeichensatzwahl, zusätzliche Attribute bei BODY und HR, Style-Sheets, ...

Informationen über HTML:

- „*Uniform Resource Locators (URL)*“; RFC 1738; Dezember 1994;
<http://rfc.fh-koeln.de/rfc/html/rfc1738.html>
- „*Hypertext Markup Language - 2.0*“; RFC 1866; November 1995;
<http://rfc.fh-koeln.de/rfc/html/rfc1866.html>
- „*HTML Tables*“; RFC 1942; Mai 1996;
<http://rfc.fh-koeln.de/rfc/html/rfc1942.html>
- „*RFC 1866 (HTML 2.0 Strict Level 1) conformance*“; 12.12.1996;
<http://www.hut.fi/~oahvenla/HTML/rfc.html>
- **Raggett, Dave:** „*Hypertext Markup Language Specification Version 3.0*“; 18.03.1995; <http://www.w3.org/pub/WWW/MarkUp/html3/html3.txt>
- „*The W3 Consortium Announces HTML 3.2 - providing bright future for HTML as a non-proprietary standard*“; <http://www.w3.org/pub/WWW/MarkUp/Wilbur/pr7may96.html>
- „*Introducing HTML 3.2*“;
<http://www.w3.org/pub/WWW/MarkUp/Wilbur/>
- **Raggett, Dave:** „*W3C Document Type Definition for the Hypertext Markup Language version 3.2*“;
<http://www.w3.org/pub/WWW/MarkUp/Wilbur/HTML32.dtd>
- **Raggett, Dave:** „*HTML 3.2 Reference Specification*“; 14.01.1996;
<http://www.w3.org/pub/WWW/TR/REC-html32.html>
- **Hubert Partl:** „*HTML-Einführung. Hypertext Markup Language die Sprache des World-Wide Web*“; August 1996;
<ftp://mail.boku.ac.at/www/htmleinf.ps.gz>

HTML-Dokumente – eine neue Dokumentenart

- HTML-Dokumente nicht vorrangig für das Ausdrucken konzipiert, sondern für die Darstellung in Bildschirmfenstern
- kein Blocksatz; aber: Dokument kann sehr leicht verschiedenen Textbreiten angepaßt werden
- keine Silbentrennung
- Absätze müssen definitiv voneinander abgegrenzt werden
- kein Seitenumbruch, keine Seiteneinteilung, keine Seitennummern
- kein Inhaltsverzeichnis (Schlagwortverzeichnis, Tabellenverzeichnis, Abbildungsverzeichnis) mit Seitennummern
- keine Fußnoten, Randnotizen, Kopf- oder Fußzeilen
- keine automatische Numerierung von Überschriften, Tabellen, Abbildungen
- vereinfachte Darstellung von Anführungszeichen, Gedankenstrichen, Auslassungspunkten
- keine automatischen Ligaturen
- anklickbare Verweise zu lokalen bzw. externen Datenquellen
- interaktive Formulare möglich
- logische Auszeichnung von Textteilen vom Autor gefordert; tatsächliche Darstellung Sache des HTML-Browsers

Vergleichende Darstellung von LaTeX und HTML

Allgemeines

The Name of the Game

TeX, LaTeX209, LaTeX2e	HTML 2.0, HTML 2.1, HTML 2.2, HTML 3.0, HTML 3.1, HTML 3.2
------------------------	--

Eingabedatei

Leerzeichen:

HTML und LaTeX: Leerzeichen und Zeilenenden einheitlich als Leerzeichen behandelt; mehrere Leerzeichen wie ein Leerzeichen

Leerzeilen:

Leerzeile bedeutet Ende eines Absatzes; mehrere Leerzeilen wie eine Leerzeile	keine besondere Bedeutung
---	---------------------------

spezielle Zeichen für die jeweilige Beschreibungssprache:

\$ & % # _ { } ~ ^ " \ < >	< > &
um sie auszudrucken, z.B.: \\$ \& \% \# _ \{ \}	um sie am Bildschirm darzustellen: das jeweilige Entity verwenden, z.B. < > &

Befehle:

eingeleitet durch \; danach entweder ein Sonderzeichen oder alphabetische Zeichen (bis zum ersten nicht-aphabetischen Zeichen)	<ul style="list-style-type: none">• Tag: eingeleitet durch <, danach das eigentliche HTML-Element, abgeschlossen durch >• Es gibt Befehle, die aus einem Anfang-Tag und Ende-Tag bestehen (Container), z.B. text, <P ALIGN="right">text</P>• In vielen Fällen kann das Ende-Tag weggelassen werden (wenn das im Kontext eindeutig ist)• Es gibt ein allgemeines Ende-Tag </> (wenn das im Kontext eindeutig ist)
LaTeX-Befehle können optionale (in eckigen Klammern) bzw. zwingende (in geschweiften Klammern) Parameter besitzen;	zusätzliche Attribute (Schlüsselwort-Parameter innerhalb der spitzen Klammern), z.B. <P ALIGN="center">
Groß-/Kleinschreibung wird unterschieden	Groß-/Kleinschreibung wird nicht unterschieden

Kommentare:

alles, was hinter dem Zeichen % folgt (bis einschließlich des Zeilenendes), ist Kommentar und wird ignoriert	Kommentare werden durch <!-- und --> eingeschlossen; können mehrere Zeilen umfassen; z.B. <!-- Das ist ein Kommentar -->
--	--

Aufbau eines Dokuments:

<code>\documentclass[optionen] %</code> <code>{dokumentklasse}</code> ... <i>Präambel mit Vereinbarungen</i> ... <code>\begin{document}</code> ... <i>eigentliches LaTeX-Dokument mit dem</i> <i>auszugebenden Text</i> ... <code>\end{document}</code>	<code><!DOCTYPE ...></code> <code><HTML></code> <code><HEAD></code> ... <code><TITLE></code> <i>Titel des HTML-Dokuments</i> <code></TITLE></code> ... <code></HEAD></code> <code><BODY></code> ... <i>eigentliches HTML-Dokument</i> <i>mit dem auszugebenden Text</i> ... <code></BODY></code> <code></HTML></code>
---	--

Erstellen der Eingabedatei:

in beiden Fällen mit einem normalen Editor; bei HTML auch mit speziellen Editoren (z.B. HoTMetaL Pro) oder durch Konverter in Textverarbeitungsprogrammen (z.B. Word Internet Assistant)

Verarbeiten der Eingabedatei:

Aufruf von LaTeX gefolgt von einem Aufruf eines DVI-Treibers oder Previewers	Laden der HTML-Datei in einem HTML-Klienten (HTML-Browser)
--	--

Setzen bzw. Ausgabe von Text

Zeilen- und Seitenumbruch

Blocksatz:

voreingestellt; kann durch verschiedene Parameter (wie z.B. <code>\parindent</code> , <code>\parskip</code>) gesteuert werden	nicht verfügbar
--	-----------------

Silbentrennung und Zeilenumbruch:

automatischer Zeilenumbruch mit Silbentrennung voreingestellt	automatischer Zeilenumbruch (ohne Silbentrennung) voreingestellt
kann durch verschiedene Parameter gesteuert werden (z.B. <code>\</code> , <code>\newline</code> , <code>\-</code> , <code>\hyphenation{muster}</code> , <code>\mbox{text}, ~</code>)	nur wenige Steuerungsmöglichkeiten (<code>
</code> , <code>&nbsp;</code> ;))

Seitenumbruch:

automatischer Seitenumbruch voreingestellt; Steuerungsmöglichkeiten durch <code>\newpage</code> , <code>\pagebreak</code> , <code>\nopagebreak</code> , <code>\enlargethispage{maß}</code>	nicht verfügbar; wird meistens durch eine Querlinie angedeutet (<code><HR></code>)
--	--

Spezielle Zeichen (Anführungszeichen, Binde- und Gedankenstriche, Auslassungspunkte, Ligaturen, Akzente und spezielle Buchstaben):

Anführungszeichen: <code>`text`</code> , <code>"text"</code> , usw.	Anführungszeichen: <code>"text"</code>
Binde- und Gedankenstriche: <code>-</code> , <code>--</code> , <code>---</code>	Binde- und Gedankenstriche: nur <code>-</code> , <code>--</code>
Auslassungspunkte: <code>\ldots</code>	Auslassungspunkte: <code>. . .</code>
automatische Ligaturen: bei <code>ff</code> , <code>fi</code> , <code>fl</code> , usw.	automatische Ligaturen: nicht verfügbar
Akzente: <code>\^{zeichen}</code> , <code>\`{zeichen}</code> usw.	Akzente: nicht verfügbar; Zeichen direkt auswählen
spezielle Buchstaben: <code>\OE</code> , <code>\ae</code> , <code>\ss</code> , usw.	spezielle Buchstaben: <code>&Oelig</code> ; <code>&aelig</code> ; <code>&szlig</code> ;

Abstände zwischen Wörtern:

<code>\frenchspacing</code> , <code>\nofrenchspacing</code> , <code>\ , \@, \, ~</code>	nur <code>nbsp</code> ;
--	-------------------------

Kapitel und Überschriften:

<code>\part{text}</code> , <code>\section{text}</code> , <code>\subsection{text}</code> , ...	<code><H1>text</H1></code> , <code><H2>text</H2></code> , ...
---	---

Fußnoten:

<code>\footnote{text}</code>	nicht verfügbar
------------------------------	-----------------

hervorgehobene Textteile:

<code>\emph{text}</code> oder <code>{\em text}</code> (ggf. mit <code>\/</code>)	leichte Hervorhebung: <code>text</code>
	starke Hervorhebung: <code>text</code>

Andere logische Textauszeichnungen in HTML:

- Zitat: `<CITE>text</CITE>`
- Code-Beispiel: `<CODE>text</CODE>`
- Benutzereingabe: `<KBD>text</KBD>`
- Folge von literalen Zeichen: `<SAMP>text</SAMP>`
- Variablennamen: `<VAR>text</VAR>`

Behandlung spezieller Textteile (Zitate, Listen, Textausrichtung, direkte Ausgabe, Abbildungen, Tabellen)

Zitate:

quote, quotation- und verse-Umgebung	<code><BLOCKQUOTE>text</BLOCKQUOTE></code>
--------------------------------------	--

Listen:

einfache Listen: <code>itemize</code> -Umgebung	einfache Listen: <code>liste</code>
numerierte Aufzählungen: <code>enumerate</code> -Umgebung	numerierte Aufzählungen: <code>liste</code>
Beschreibungen: <code>description</code> -Umgebung	Beschreibungen: <code><DL>liste</DL></code>
Listeneinträge jeweils durch <code>\item</code> bzw. <code>\item[term]</code> eingeleitet	Listeneinträge jeweils durch <code></code> bzw. <code><DT>term<DD></code> eingeleitet

Textausrichtung:

Blocksatz voreingestellt	Blocksatz nicht verfügbar
andere Ausrichtungen durch <code>flushleft</code> -, <code>flushleft</code> - oder <code>center</code> -Umgebung	andere Ausrichtungen durch <code><CENTER>text</CENTER></code> oder das Attribut <code>ALIGN</code> bei den Elementen <code>CAPTION</code> , <code>DIV</code> , <code>INPUT</code> , <code>P</code> , <code>TABLE</code> , <code>TD</code> , <code>TH</code> , <code>TR</code> (jeweils als Container formuliert); mögliche Attribut-Werte: <code>center</code> , <code>left</code> , <code>right</code> ; z.B. <code><P ALIGN=center>text</P></code>

direkte Ausgabe:

<code>\verb text </code>	<code><SAMP>text</SAMP></code> oder <code><CODE>text</CODE></code>
verbatim-Umgebung	beste Entsprechung durch <code><XMP>text</XMP></code> (<code><LISTING>text</LISTING></code>) <code><PRE>text</PRE></code> läßt Zeichenfor- matierungen und Konstruktionen der Art <code><A attribute>text</code> zu

Abbildungen:

PostScript-Bilder können mit Hilfe zu- sätzlicher Pakete eingebaut werden.	Inline-Bilder: <code></code> anklickbare Verweise auf externe Bilder durch <code></code> <code>anker</code>
Bildunterschrift durch die Anweisung <code>\caption{text}</code> in einer figure- Umgebung	keine Bildunterschrift in dem Sinne verfügbar

Tabellen:

Tabellierungen: tabbing-Umgebung	Tabellierungen: nicht verfügbar
eigentliche Tabellen: tabular-Umgebung	eigentliche Tabellen: <TABLE> <i>tabelle</i> </TABLE>
Tabellenüberschrift durch die Anweisung \caption{ <i>text</i> } in einer table-Umgebung	Tabellentitel durch <CAPTION> <i>text</i> </CAPTION>
Beispiel für eigentliche Tabelle: \begin{table} \caption{Kleine Tabelle} \begin{tabular}{ l c r} \hline zelle 1 & zelle 2 & zelle 3 \\ zelle 4 & zelle 5 & zelle 6 \\ \hline \end{tabular} \end{table}	Beispiele für (einfache) Tabelle: <TABLE BORDER> <CAPTION> Kleine Tabelle </CAPTION> <TR><TD>zelle 1 <TD>zelle 2 <TD>zelle 3 <TR><TD>zelle 4 <TD>zelle 5 <TD>zelle 6 </TABLE>

Querverweise innerhalb des Dokuments:

Verweisziel: \label{ <i>label</i> }	Verweisziel: <A NAME=" <i>name</i> " optionen> <i>text</i>
Verweis: \ref{ <i>label</i> } bzw. \pageref{ <i>label</i> }	Verweis: <A HREF="# <i>name</i> " optionen> <i>anker</i>

Spezialitäten

Schriftarten und Schriftgrößen

normal: <code>\textrm{text}</code> bzw. <code>{\rm text}</code>	normal:
Sans-Serif: <code>\textsf{text}</code> bzw. <code>{\sf text}</code>	Sans-Serif: nicht direkt verfügbar
Typewriter: <code>\texttt{text}</code> bzw. <code>{\tt text}</code>	Typewriter: <code><TT>text</TT></code>
fett: <code>\textbf{text}</code> bzw. <code>{\bf text}</code>	fett: <code>text</code>
kursiv: <code>\textit{text}</code> bzw. <code>{\it text}</code>	kursiv: <code><I>text</I></code>
nicht direkt verfügbar	durchgestrichen: <code><STRIKE>text</STRIKE></code>
schräg: <code>\textsl{text}</code> bzw. <code>{\sl text}</code>	schräg: nicht direkt verfügbar
Kapitalchen: <code>\textsc{text}</code> bzw. <code>{\sc text}</code>	Kapitalchen: nicht direkt verfügbar
Schriftgrößen: <code>\tiny</code> , ..., <code>\small</code> , <code>\normalsize</code> , <code>\large</code> , ..., <code>\Huge</code>	Schriftgrößen: <code>text</code> , wobei <i>größe</i> eine konkrete Zahl (mit oder ohne Vorzeichen) ist

Abstände in LaTeX

(in HTML nicht verfügbar)

- Zeilenabstand: `\linespread{faktor}`
- spez. horizontale Abstände: `\,`, `\enspace`, `\quad`, `\qquad`, `\hfill`
- spez. vertikale Abstände: `\vspace{länge}`, `\smallskip`, `\medskip`, `\bigskip`, `\vfill`

Style-Sheets

- HTML ursprünglich als Dokumentenbeschreibungssprache konzipiert.
- Einführung vieler Formatierungselemente und -attribute auf niedriger Ebene verwischt diese Zielsetzung
- Style-Sheets ermöglichen dem Autor, Inhalt und Darstellung des Textes wieder zu trennen.
- Style-Sheet kann zwischen `<STYLE>...</STYLE>` im Kopf des Dokuments vereinbart werden.
- Autor kann sich auch auf externe Style-Sheets beziehen.
- eine sehr einfache Form eines Style-Sheets:

```
P      {font-size : 12pt}  
STRONG {color : #800000}
```

→ „normale“ Absätze: in der Schriftgröße 12pt,
→ „starke“ Hervorhebungen: dunkelrot

- noch zahlreiche andere Mechanismen verfügbar

„Übriggebliebenes“

In LaTeX: Formeln, kompliziertere Tabellen

- Außer `_{text}` und `^{text}` sind in HTML 3.2 keine weiteren Konstruktionen für Formeln verfügbar.
- Im Vergleich zu HTML 3.2 können in LaTeX deutlich komplexere Tabellen erstellt werden.

In HTML: Hypertext-Strukturen, Formulare

Hypertext-Strukturen:

Mit Hilfe der Konstruktion `anker` kann in HTML auf lokale oder externe Datenquellen verwiesen werden.

URL (Uniform Resource Locator) hat hierbei den folgenden Aufbau:

protokoll : // *rechner* / *pfad*

mit

protokoll Protokoll oder Zugriffsmechanismus auf *rechner*; folgende Angaben sind u.a. möglich (http, mailto, ftp, news, telnet)

rechner Internet-Adresse der Datenquelle (ggf. mit Port-Nummer)

pfad führt auf *rechner* zur gewünschten Datei

Beispiele:

`http://www.uni-giessen.de/hrz/texservice.html`

`ftp://ftp.uni-giessen.de/pub/doc/gopher.faq`

`news:de.comp.text.tex`

`mailto:Guenter.Partosch@hrz.uni-giessen.de`

Formulare:

```
...
<FORM
ACTION="http://www.uni-giessen.de/cgi-bin/xyz.gi"
METHOD="POST">

<H1>Anmeldung zur Tagung DANTE'98</H1>

<P>
Name  :<INPUT NAME="Text1" VALUE="Partosch" >
<P>
Vorname  :<INPUT NAME="Text2" VALUE="Günter" >
<P>
Ich werde
<P>
<INPUT TYPE="CHECKBOX" NAME="Kontrollkästchen1"
CHECKED> am Er&ouml;ffnungsvortrag,
<P>
<INPUT TYPE="CHECKBOX" NAME="Kontrollkästchen2"
CHECKED> an den Votr&auml;gen am Vormittag,
<P>
<INPUT TYPE="CHECKBOX" NAME="Kontrollkästchen3" > am
gemeinsamen Mittagessen,
<P>
<INPUT TYPE="CHECKBOX" NAME="Kontrollkästchen4"
CHECKED> an den Votr&auml;gen am Nachmittag und
<P>
<INPUT TYPE="CHECKBOX" NAME="Kontrollkästchen5"
CHECKED> am Abschlus&szlig;vortrag
<P>
teilnehmen.
<P>
Die Teilnahmegeb&uuml;hr bezahle ich :
<SELECT NAME="Dropdown1" >
<OPTION>per Kreditkarte
<OPTION>per &Uuml;berweisung
<OPTION>per Lastschrift
<OPTION>per Euro-Scheck
<OPTION>bar</SELECT>
<P>
```

```
<INPUT TYPE="SUBMIT" VALUE="Abschicken">
<P>
<INPUT TYPE="RESET">
</FORM>
...
```