

Grundbegriffe der Informatik

Kapitel 11: Dokumente

Thomas Worsch

KIT, Institut für Theoretische Informatik

Wintersemester 2015/2016

Wo sind wir?

Dokumente

XHTML

\LaTeX

Eine Grenze unserer bisherigen Vorgehensweise

Dokumente haben Inhalt, Struktur und Form

überall Inschriften

- Briefe, Kochrezepte, Zeitungsartikel, Vorlesungsskripte
- Seiten im WWW, Emails, ...
- diese Folien

drei Aspekte, die für Leser relevant

- **Inhalt** des Textes
- **Struktur** des Textes
- **Erscheinungsbild**, (äußere) **Form**

Inhalt, Struktur und Form

- Inhalt des Textes
- Struktur des Textes
- Erscheinungsbild, (äußere) Form

Inhalt, Struktur und Form

- Inhalt des Textes
- Struktur des Textes
- Erscheinungsbild, (äußere) Form

andere Form:

- INHALT des Textes
- STRUKTUR des Textes
- ERSCHEINUNGSBILD, (äußere) FORM

Inhalt, Struktur und Form

- Inhalt des Textes
- Struktur des Textes
- Erscheinungsbild, (äußere) Form

andere Struktur:

Inhalt des Textes, Struktur des Textes,
Erscheinungsbild, (äußere) Form

Inhalt, Struktur und Form

- Inhalt des Textes
- Struktur des Textes
- Erscheinungsbild, (äußere) Form

anderer Inhalt:

- *Balaenoptera musculus* (Blauwal),
- *Mesoplodon carlhubbsi* (Hubbs-Schnabelwal) und
- *Physeter macrocephalus* (Pottwal).

Wozu Inhalt, Struktur und Form?

Inhalt für Autoren und Leser im Vordergrund

- (Ausnahmen?)

Struktur und Form

- Unterstützung beim Verstehen des Inhalts

Wozu Inhalt, Struktur und Form?

Inhalt für Autoren und Leser im Vordergrund

- (Ausnahmen?)

Struktur und Form

- Unterstützung beim Verstehen des Inhalts

Dokumente: Texte mit Inhalt, Struktur und Form

- z. B. Lösungen für Hausaufgaben, Programme, ...

Wozu Inhalt, Struktur und Form?

Inhalt für Autoren und Leser im Vordergrund

- (Ausnahmen?)

Struktur und Form

- Unterstützung beim Verstehen des Inhalts

Dokumente: Texte mit Inhalt, Struktur und Form

- z. B. Lösungen für Hausaufgaben, Programme, ...

ein **Rat für Ihre vielen Dokumente**

- Finden geeigneter Struktur hilft auch eigenem Verständnis
- *früh damit beginnen, etwas aufzuschreiben*
(weil man dann über Struktur nachdenken muss)
- und \LaTeX benutzen

Struktur von Dokumenten

syntaktische Korrektheit spielt eine Rolle

- zumindest wenn Rechner im Spiel sind

Beispiele aus **Auszeichnungssprachen**
(engl. *markup language*)

- Tabellen in XHTML
- Listen in \LaTeX
 - und ein klein bisschen Allgemeines zu \LaTeX

Wo sind wir?

Dokumente

XHTML

L^AT_EX

Eine Grenze unserer bisherigen Vorgehensweise

XHTML

HTML: Auszeichnungssprache für WWW-Seiten

XHTML: sozusagen noch striktere Variante von HTML

für beide formaler als für \LaTeX festgelegt, wie syntaktisch korrekte solche Seiten aussehen.

durch sogenannte **document type definition**, *DTD*.

Auszug aus der DTD für Tabellen in XHTML

```
<!ELEMENT table      (caption?, thead?, tfoot?, (tbody+|tr+))>
<!ELEMENT caption   %Inline;>
<!ELEMENT thead     (tr)+>
<!ELEMENT tfoot     (tr)+>
<!ELEMENT tbody    (tr)+>
<!ELEMENT tr        (th|td)+>
<!ELEMENT th        %Flow;>
<!ELEMENT td        %Flow;>
```

Interpretation der DTD für Tabellen in XHTML

```
<!ELEMENT table (caption?, thead?, tfoot?, (tbody+|tr+))>
<!ELEMENT thead (tr)+>
<!ELEMENT tfoot (tr)+>
<!ELEMENT tbody (tr)+>
<!ELEMENT tr (th|td)+>
```

`table`, `thead`, `tr`: Namen für formale Sprachen

Bedeutung von `+` ϵ -freier Konkatenationsabschluss

Bedeutung von `,` Produkt formaler Sprachen

Bedeutung von `|` Vereinigung

Bedeutung von `?` $L^? = L^0 \cup L^1 = \{\epsilon\} \cup L$

- „optionales“ Auftreten eines Wortes aus L

Interpretation der DTD für Tabellen in XHTML

`<!ELEMENT tbody (tr)+ >` legt fest:

$$L_{tbody} = \{<tbody>\} \cdot L_{tr}^+ \cdot \{</tbody>\}$$

Tabellentrumpf beginnt mit `<tbody>`, endet mit `</tbody>`,
dazwischen positive Anzahl von Tabellenzeilen

erste Zeile aus der DTD besagt:

$$L_{table} = \{<table>\} \cdot L_{caption}^? \cdot L_{thead}^? \cdot L_{tfoot}^? \cdot L_{tbody}^+ \cdot \{</table>\}$$

Tabelle von `<table>` und `</table>` umschlossen und enthält

- optional eine Überschrift (*caption*),
- optional einen Tabellenkopf (*table head*),
- optional einen Tabellenfuß (*table foot*) und
- eine beliebige positive Anzahl von Tabellentrümpfen.

Beispiel für Tabelle in XHTML

```
<table>
  <tbody>
    <tr> <td>1</td> <td>a</td> </tr>
    <tr> <td>2</td> <td>b</td> </tr>
  </tbody>
</table>
```

Wo sind wir?

Dokumente

XHTML

\LaTeX

Eine Grenze unserer bisherigen Vorgehensweise

basiert auf T_EX von Donald Knuth

<http://www-cs-faculty.stanford.edu/~knuth/>

Textsatz-Programm

in der Informatik sehr häufig verwendet

- hervorragender automatischer Satz mathematischer Formeln
- aus
 - \[2 - \sum_{i=0}^k i 2^{-i} = (k+2) 2^{-k} \]
- wird

$$2 - \sum_{i=0}^k i 2^{-i} = (k + 2) 2^{-k}$$

Skript, Folien, etc. von GBI mit L^AT_EX gemacht

L^AT_EX (2)

in L^AT_EX-Quelle des Skriptes

```
\section{Struktur von Dokumenten}
```

woraus L^AT_EX die Zeile

11.2 STRUKTUR VON DOKUMENTEN

im Skript gemacht hat

- automatisch passende Abschnittsnummer eingefügt
- alles in Großbuchstaben gesetzt

z. B. Schriftauswahl *nicht* in der Eingabe vermerkt

Diese Festlegung an anderer Stelle, und zwar

- an *einer* Stelle das typografische Aussehen
- *aller* Abschnittsüberschriften (einheitlich) festgelegt

Grobstruktur von \LaTeX -Dokumenten

```
\documentclass[11pt]{report}
% so schreibt man Kommentare
% dieser Teil heißt Präambel des Dokumentes
  \usepackage[T1]{fontenc}      % ohne Erklärung ....
  \usepackage[ngerman]{babel}  % deutsche Trennungen etc.
  \usepackage[utf8]{inputenc}  % Zeichensatz der Eingabe

  \author{Thomas Worsch}
  \title{Dokumente mit LaTeX}
\begin{document}
% und hier der eigentliche Text .....
  \maketitle
  \chapter{Überschrift}
   $a^2 + b^2 = c^2$       % eine kleine Formel
\end{document}
```

Listen mit \LaTeX

einfache Aufzählungsliste in \LaTeX

Eingabe	Ausgabe
<code>\begin{itemize}</code>	
<code>\item Inhalt</code>	• Inhalt
<code>\item Struktur</code>	• Struktur
<code>\item Form</code>	• Form
<code>\end{itemize}</code>	

Punkt als Markierung *nicht in* der Liste festgelegt

Trennung der Spezifikationen von **Struktur** und **Form**

„–“ statt „•“ erfordert *eine* Änderung in Präambel

Formale Sprachen kommen ins Spiel

Gesucht: formale Sprache L_{itemize}

- aller legalen Texte für Listen in LaTeX

Gegeben (tun wir mal so ...)

- formale Sprache L_{item} aller Texte,
- die hinter einem Aufzählungspunkt vorkommen dürfen

Dann

$$L_{\text{itemize}} = \{\backslash\text{begin}\{\text{itemize}\}\} \left(\{\backslash\text{item}\} L_{\text{item}} \right)^+ \{\backslash\text{end}\{\text{itemize}\}\}$$

Problem: in Aufzählungspunkten wieder Listen erlaubt ...

$$L_{\text{item}} = \dots L_{\text{itemize}} \dots$$

$$L_{\text{itemize}} = \dots L_{\text{item}} \dots$$

Gleichungssystem lösbar? eindeutig lösbar?

Wo sind wir?

Dokumente

Eine Grenze unserer bisherigen Vorgehensweise

Eine Grenze unserer bisherigen Vorgehensweise

eben mit Hilfe von Produkt und Konkatenationsabschluss formaler Sprachen präzise Aussagen gemacht.

Das ging, weil

- etwas von einer komplizierteren Art aus Bestandteilen einfacherer Art zusammengesetzt wurde.

Eine Grenze unserer bisherigen Vorgehensweise

eben mit Hilfe von Produkt und Konkatenationsabschluss formaler Sprachen präzise Aussagen gemacht.

Das ging, weil

- etwas von einer komplizierteren Art aus Bestandteilen einfacherer Art zusammengesetzt wurde.

manchmal

- größere Dinge einer Art zusammengesetzt aus kleineren Bestandteilen *der gleichen Art*
- z. B. Listen, deren Aufzählungspunkte wieder Listen enthalten

Eine Grenze unserer bisherigen Vorgehensweise (2)

anderes typisches Beispiel: korrekte Klammerungen
(wie bei arithmetischen Ausdrücken)

zu jeder *Klammer auf* „weiter hinten“ die „zugehörige“
Klammer zu

insbesondere legal:

- mehrere korrekte Klammerungen hintereinander
- zusätzliche Klammern um korrekte Klammerung außen herum

schön wäre z. B. Beziehung von

L_{Klammer} mit L_{Klammer}^* und mit $\{() \cdot L_{\text{Klammer}} \cdot ()\}$

aber wie? Gleichung?

- lösbar?
- wenn ja: eindeutig?
- ein Thema des nächsten Kapitels ...

Was ist wichtig

Das sollten Sie mitnehmen:

- Dokumente haben **Inhalt**, **Struktur** und **Form**.
- formale Sprachen helfen bei der Definition legaler Strukturen
- früh mit dem Aufschreiben beginnen

Das sollten Sie üben:

- Dokumente mit **L^AT_EX** aufschreiben