



openHPI HTML-Dokumente

Prof. Dr. Christoph Meinel

Hasso-Plattner-Institut

Universität Potsdam

Erinnerung

Erfolg des WWW – World Wide Web – basiert im Wesentlichen auf seinen Inhalten, den verlinkten **Hypermedia-Dokumenten**

HTML – Hypertext Markup Language – ist die einfache Markupsprache zur Beschreibung dieser Dokumente

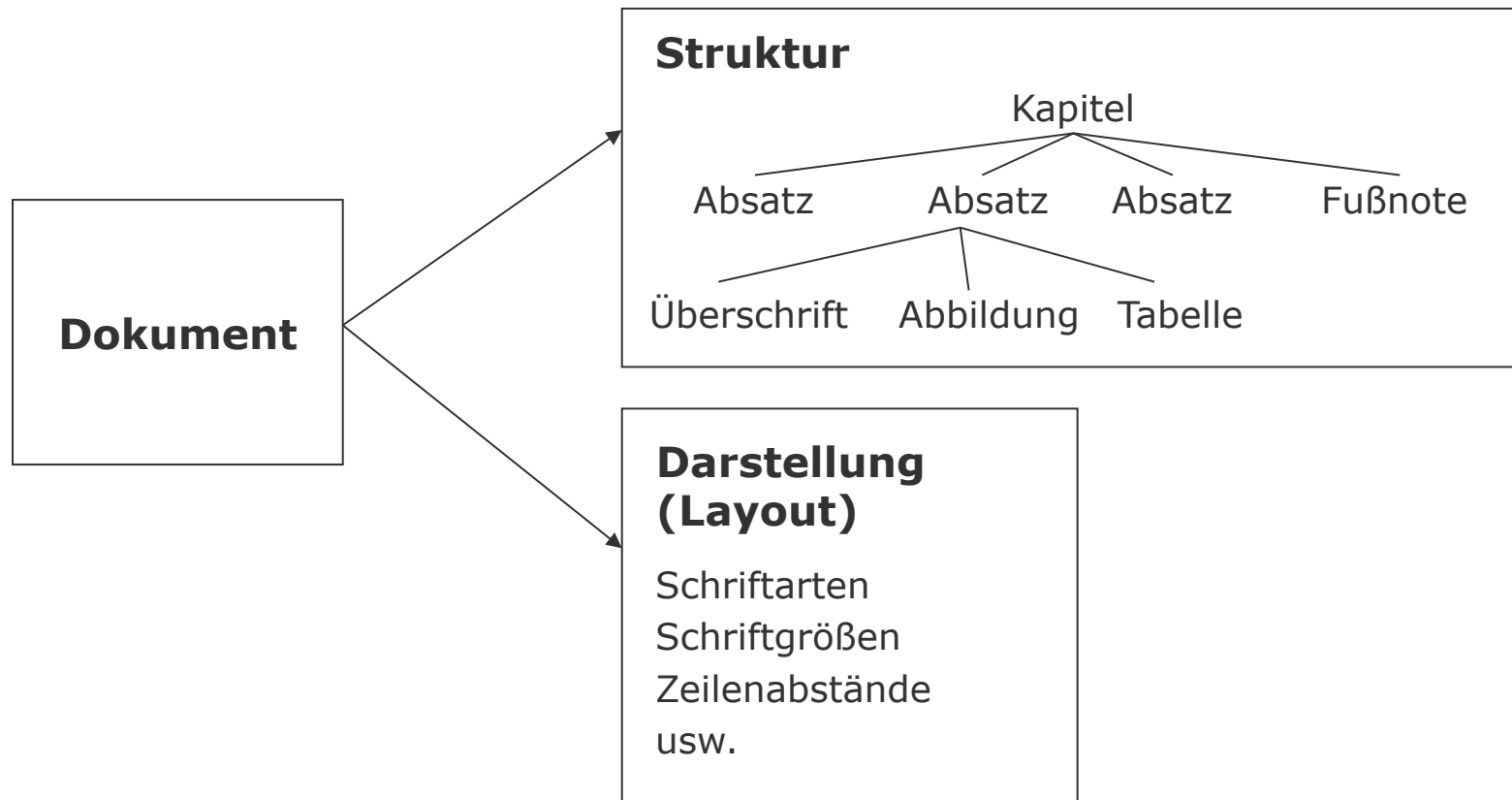
- HTML ist neben URIs und HTTP dritte Basiskomponente des WWW
- HTML gibt weder Format noch Layout eines Hypermedia-Dokuments vor, sondern beschreibt im Wesentlichen die **Struktur** der jeweiligen Inhalte
- Einfache Beschaffenheit von HTML ist wichtiger Grund für den Erfolg des WWW
- HTML baut ursprünglich auf der komplexeren Markupsprache **SGML – Standard Generalized Markup Language** auf, deren Prinzipien noch heute gelten

SGML (1/3)

SGML – Standard Generalized Markup Language – wurde in den 1960er entwickelt, als moderne Datenverarbeitung ihren Anfang nahm

- SGML sollte den Austausch strukturierter Daten ermöglichen, unabhängig von Herstellern, Geräten oder Plattformen
- SGML ist eine Meta-Markupsprache, in der sich unterschiedliche Datenformate (genauer: Markupsprachen für Datenformate) für alle möglichen (Text-)Dokumente beschreiben lassen
- **Grundlegende Idee:** Trennung von
 - **Dokumentstruktur** – logische Unterteilung und Abhängigkeiten der einzelnen Teile eines Dokuments – und
 - **Dokumentendarstellung** – Formatierung und Layout
- **Vorteil:** Aus einem SGML-Dokument könnten viele verschiedene Layout-Varianten generiert werden, z.B. für Druck im Magazin oder als Buch, für die Anzeige auf Computer oder Smartphone, ...

Trennung zwischen Struktur und Darstellung



SGML (3/3)

- Strukturgebende Elemente werden mittels sogenannter **Markups** mit unterschiedlichen **Tags** gebildet
 - Markups sind durch Kombinationen spezieller Zeichen gekennzeichnet und werden innerhalb des eigentlichen Dokuments gespeichert
 - Markup-Begrenzer – „<...>“, z.B. <kapitel> ... </kapitel>
 - Interpreter können so die Tags leicht als solche erkennen
- Beziehungen zwischen den strukturellen Elementen eines SGML-Dokumentes werden durch den SGML **Dokumentenbaum** beschrieben
- SGML erlaubt als Metasprache die Definition eigener Dokumentklassen, z.B. für ein „Buch“, um Beschreibung von Instanzen dieser Klasse zu vereinheitlichen und **automatische Syntaxprüfung** zu ermöglichen
 - Dokumentenklassen werden mittels von **DTDs – Document Type Definitions** – festgelegt

- **HTML** ist eine Markupsprache, die auf SGML basiert
- HTML ist **einfacher** als SGML, da es keine Meta-Markupsprache ist – HTML ist eine Sprache, die „in SGML formuliert“ ist ...
- HTML führt das Hyperlink-Konzept des WWW ein
- **HTML-Dokumente** werden im Browser dargestellt; dieser muss dazu einen HTML-Interpreter besitzen, der die einzelnen strukturgebenden Elemente im Dokument identifizieren und für die sachgerechte Anzeige vorbereiten kann

- Historisch bedingt wird bei HTML die Trennung von Struktur und Darstellung nicht strikt verwirklicht, die SGML eigentlich fordert, sondern HTML bietet auch Reihe von Möglichkeiten zur Formatierung des Inhalts
- Dennoch können (und sollten) Autoren in HTML keine direkten Formatierungsanweisungen verwenden und stattdessen die Darstellung des Dokuments mittels separater Formatierungsanweisungen beschreiben
→ **Cascading Stylesheets (CSS)**
- Viele der ursprünglichen Formatierungsmöglichkeiten in HTML gelten heute als „ausgemustert“; insbesondere mit Einführung von HTML5 wurde eine ganze Reihe dieser alten Elemente gestrichen

- Ein **HTML-Dokument** beginnt immer mit der Angabe der verwendeten HTML-Version, z.B.
 - `<DOCTYPE html>` für HTML5-Dokumente
- Ein HTML-Dokument besteht immer aus einem **Header** und einem **Body**
- Der **Header** enthält **Informationen über das Dokument**, die nicht zum eigentlichen Inhalt gehören, z.B.
 - Titel, Autor, Keywords, Kodierung, Sprache, ...
- Der **Body** enthält **eigentlichen Inhalt** des Dokuments, z.B.
 - Überschriften, Absätze, Listen, Tabellen, Formulare, Referenzen zu Bildern, Hyperlinks, usw.

Grundgerüst eines HTML-Dokuments

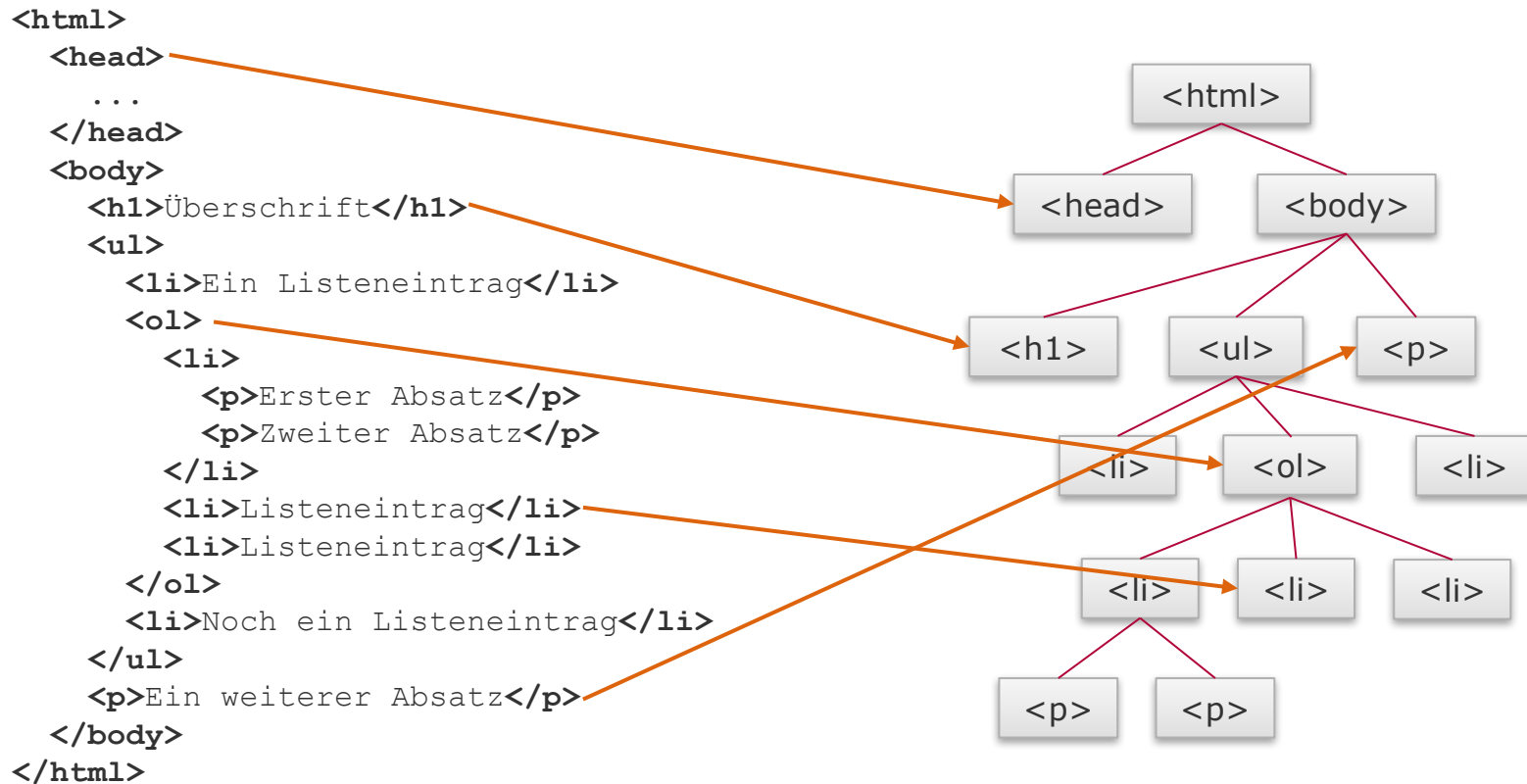
```
<!DOCTYPE html>
<html>
  <head>
    <title>Über openHPI</title>
    ...
  </head>
  <body>
    <h1>openHPI - die MOOC-Plattform des HPI</h1>
    <p>MOOCs - Massive Open Online Courses - bieten
      eine der innovativsten Lernformen: Frei über
      das Internet können Lernende auf multimediale
      Kursmaterialien zugreifen...</p>
    ...
  </body>
</html>
```

Header

Body

- HTML-Elemente werden – wie in SGML – durch **Markup** deklariert
- Markup wird durch Markup-Bezeichner – sogenannte **Tags** – gekennzeichnet, die jeweils von den Markup-Begrenzern **<** und **>** umschlossen werden
- Strukturelle Elemente des HTML-Dokuments werden jeweils von einem **Start-Tag** und einem **End-Tag** eingefasst, z.B.
 - `<p> Absatz </p>` (engl. „Paragraph“)
- **End-Tags** können in einigen Fällen weggelassen werden
- Tags können **Attribute** für zusätzliche Informationen enthalten, z.B. `<p id="absatz1">...</p>`
- Die aktuelle HTML-Version – HTML5 – kennt etwa 120 verschiedene Tags

- HTML-Elemente werden immer sequenziell angeordnet oder ineinander verschachtelt (niemals „über Kreuz“)
➔ dadurch ergibt sich Baumstruktur: **HTML-Dokumentenbaum**



Document Object Model – DOM

- **Document Object Model – DOM** – ist standardisierte Schnittstelle für **programmatischen Zugriff** auf Elemente eines HTML-Dokuments
- Zugriff auf DOM erfolgt mit Client-seitiger Programmierung (JavaScript)
- DOM erlaubt
 - Navigation durch Dokumentenbaum („Auswählen“ von einzelnen oder mehreren HTML-Elementen)
 - Hinzufügen, Löschen oder Verschieben von HTML-Elementen
 - Auslesen oder Ändern von Element-Inhalten (Text)
 - Auslesen oder Ändern von Attributen
- Beispiele:
 - `document.getElementsByTagName("p")` – selektiert alle `<p>`-Tags
 - `document.getElementById("Liste").children` – selektiert alle Unterelemente („Kindknoten“) des Elements mit der Id „Liste“