

Die Basis des Web: HTML und HTTP

Stephan Karrer

HTML

HTML ist die Sprache, in der die Web-Seiten erstellt werden.

- HTML: Hypertext Markup Language. Erfinder: Tim Berners-Lee (1989/1990)
- Standardisiert durch das W3C (World Wide Web Consortium).
 - aktuelle Version: HTML 4.01, XHTML 1.0
 - Draft: HTML 5.0 (August 2008)
- Text wird strukturell ausgezeichnet. HTML ist somit eine Auszeichnungssprache.
- In HTML lassen sich Texte gliedern in Überschriften, Absätze, Listen und Tabellen. Innerhalb dieser Blockelemente werden Formatanweisungen für Zeichenfolgen benutzt.
- Abgespeichert werden Hypertext-Dokumente üblicherweise mit der Dateiendung «.html» (oder «.htm»).

Grundaufbau einer HTML-Seite

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
    "http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
    <title> Ersetze durch Titel </title>
</head>
<body>
    Ersetze durch Content
</body>
</html>
```

Verweise (links) in HTML

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
    "http://www.w3.org/TR/html4/loose.dtd">
<html> <head> <title> Weiterführende Informationen </title> </head>
<body> Weiterführende Infos bei
    <a href="http://www.selfhtml.org/">SELFHTML</a>
</body>
</html>
```

- HTML bietet die Möglichkeit, an eine andere Stelle innerhalb eines Dokuments oder eines anderen Dokuments zu springen.
- Diese Verknüpfungen werden gesteuert durch Hyperlinks (oder einfach Links).
- Links sind heute das wichtigste Element zum Surfen im Web.

Seitenlayout mit Frames

- Mittels Frames kann die Darstellung auf mehrere Bereiche mit jeweils eigenem Inhalt verteilt werden
 - Die Frame-Technik wird durch die meisten Browser unterstützt und ist auch durch W3C standardisiert.
 - Eigenheiten:
 - Jeder Frame hat eigene Historie
 - Adressierung ist schwieriger
 - Darstellung ist abhängig von der Bildschirmauflösung
 - Suchmaschinen tun sich schwerer
- werden zunehmend weniger eingesetzt
- heute üblicherweise Layout per CSS



Quelle: www.selfhtml.org

Seitenlayout mit CSS (Cascading Stylesheets)

- Heute übliche Technik für das Seitenlayout
 - W3C-Standard (CSS Level 1, CSS 2.1)
 - wird von den gängigen Browsern unterstützt
- Erlaubt die Trennung von Layout und Inhalt
- Layout-Möglichkeiten:
 - Textformatierung
 - Schriftformatierung (Fonts)
 - Farben und Hintergrund
 - Absätze und Elementboxen
 - Positionierung
- Verwendung mehrfacher Styles, auch geschachtelt, möglich

```
h1 { color: white;
background: orange;
border: 1px solid black;
padding: 0 0 0 0;
font-weight: bold;
}
/* begin: seaside-theme */

body {
background-color:white;
color:black;
font-family:Arial,sans-serif;
margin: 0 4px 0 0;
border: 12px solid;
}
```

CSS

Quelle: Wikimedia Commons

Interaktion via Formularen

Quelle: W3C

```
<FORM action="http://somesite.com/prog/adduser" method="post">
  <P>    <LABEL for="firstname">First name: </LABEL>
        <INPUT type="text" id="firstname"><BR>
        <LABEL for="lastname">Last name: </LABEL>
        <INPUT type="text" id="lastname"><BR>
        <LABEL for="email">email: </LABEL>
        <INPUT type="text" id="email"><BR>
        <INPUT type="radio" name="sex" value="Male"> Male<BR>
        <INPUT type="radio" name="sex" value="Female"> Female<BR>
        <INPUT type="submit" value="Send"> <INPUT type="reset">

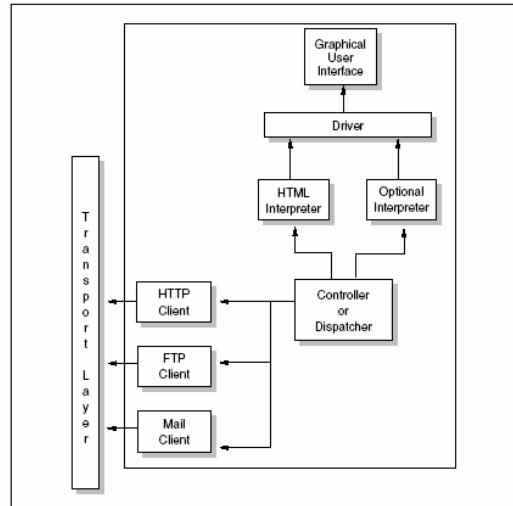
  </P>
</FORM>
```

Über Formulare können Benutzereingaben getätigt werden

- Die Eingabedaten werden als Zeichenketten zum Web-Server übertragen
- Applikationslogik auf der Serverseite verarbeitet diese und generiert in der Regel dynamisch eine entsprechende Antwortseite

Webbrowser als universeller Client

- Browser bieten häufig auch die Integration weiterer Dienste: Email, Newsgroups, ftp
- HTML erlaubt die Integration multimedialer Inhalte: Bilder, Grafiken, Audio, Video, ...
- Übliche Browser lassen sich über ein Plugin-Konzept erweitern: Flash, Java, PDF, ...



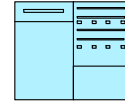
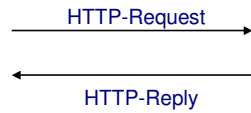
Quelle: IBM

Kommunikation via HTTP

- Standardisiert durch IETF
- Versionen: HTTP/1.0, HTTP/1.1
- Benutzt TCP
- Kommunikationsschema:
Request - Response



Browser



Webserver

```
GET /dbag/dispatch/de/kir/gdb_navigation/home
HTTP/1.1
Host: deutsche-boerse.com
User-Agent: Mozilla/5.0 (Windows; U; ...
Accept: text/xml,application/xml, ...
Accept-Language: de-de,de;q=0.8,en-us;q=0.5, ...
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7 ...
Keep-Alive: 300
Connection: keep-alive
```

```
HTTP/1.1 200 OK
Date: Sun, 03 Aug 2008 11:28:10 GMT
Cache-Control: no-store
Pragma: no-cache
Expires: Thu, 01 Jan 1970 00:00:00 GMT
Last-Modified: Sun, 03 Aug 2008 11:28:10 GMT
Content-Type: text/html;charset=UTF-8
Keep-Alive: timeout=5, max=98
Connection: Keep-Alive
Content-Language: de
```

Das HTTP-Protokoll

HTTP Methoden	
GET	Ressource-Anforderung
POST	Übertragung von Daten zum Server
HEAD	Server soll nur Header senden
...	

HTTP Stauscodes	
1xx	Zur Information
2xx	Erfolgreiche Operation
3xx	Umleitung
4xx	Fehler auf der Client-Seite
5xx	Fehler auf der Serverseite

Mime Typen		
Typ	Dateiendung	Bedeutung
application/zip	*.zip	ZIP-Archivdateien
image/jpeg	*.jpeg *.jpg *.jpe	JPEG-Dateien
text/html	*.htm *.html *.shtml	HTML-Dateien
text/plain	*.txt	reine Textdateien
...

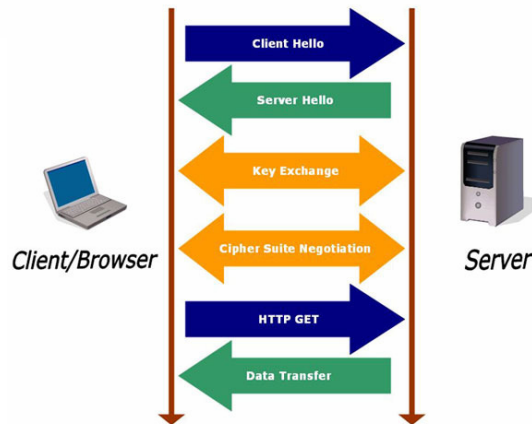
Zustandsverwaltung (Session-Management)

- HTTP ist zustandslos
 - Session-Management und Transaktionsmanagement sind Herausforderungen

Ansätze:

- Verborgene Formularfelder
 - Der Server fügt verborgene Felder in jedes gesendete Formular ein, die dann stets wieder mitgeliefert werden
- URL-Rewriting
 - Die dynamisch erzeugten HTML-Seiten enthalten URLs, die um Session-Informationen erweitert sind.
- Cookies
 - Server kann mit jeder Response-Nachricht Session-Informationen an den Client zur Speicherung übertragen.
 - Bei jedem weiteren Zugriff auf diesen Server überträgt der Client diese Information.
- Eigene Session, z.B. gesicherter Kommunikationskanal via HTTPS (HTTP over SSL)

Geschützte Kommunikation via SSL bzw. TLS



Quelle: <https://www.securetrust.com/resources/how-ssl-works>

- SSL bzw. sein Nachfolger TLS werden üblicherweise für eine abgeschirmte Kommunikation eingesetzt.
- Es können verschiedene Verschlüsselungsverfahren ausgehandelt werden.
- Bei den heute üblichen Szenarien (HTTPS) identifiziert sich nur der Server anhand eines Zertifikats.
 - Server benötigt gültiges Zertifikat
 - Client überprüft das Server-Zertifikat
- Vorsicht: nur der Transport ist Ende-zu-Ende verschlüsselt
 - Die Echtheit der Daten und deren Verarbeitung ist nicht geregelt.