

Ergänzungs-Studienangebot der Mediendidaktik für Lehramtstudenten Dr. rer. nat. Harald Sack Institut für Informatik FSU Jena

Wintersemester 2005/2006



1 2 3 4 5 6 7 8 9 10 18.01.2006 - Vorlesung Nr. 11 12

13

14

3. Internet und WWW (Teil 5)



## 3. Internet und WWW (6)

- O World Wide Web
  - Grundbegriffe
  - Browser und WWW-Server
  - eindeutige Identifikation durch URL
  - Kommunikation über HTTP
  - Kodierung mit HTML / CSS
  - O Interaktivität via
    - Dynamisches HTML
    - O cgi-Programme
    - Cookies
  - die Zukunft gehört XML

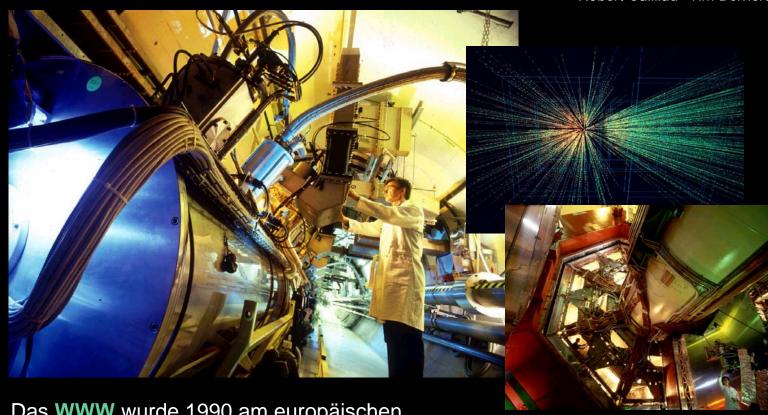
## World Wide Web





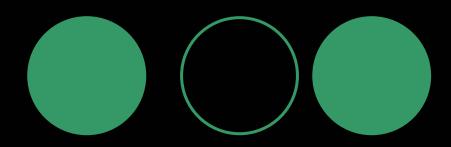
Die Geburt des World Wide Web

Robert Cailliau Tim Berners Lee



Das **WWW** wurde 1990 am europäischen Kernforschungszentrum CERN geboren



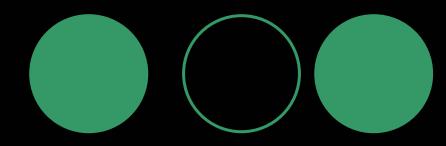


- Grundbegriffe
  - Versuch einer Definition.....

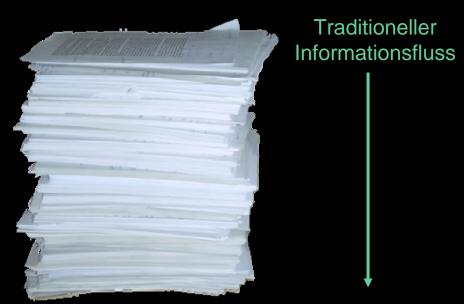
#### World Wide Web / WEB / WWW / W3

- ist ein riesiges Online-Informationslager, auf das mit Hilfe eines interaktiven Anwendungsprogrammes namens "Browsers" zugegriffen werden kann
- Internet-Ressourcen, auf die mit Hilfe des HTTP-Protokolls (Hypertext Transfer Protocol) zugegriffen werden kann
- Weltweit verteilte, multimediale Informationsressourcen, die untereinander über Hyperlinks miteinander verbunden sind

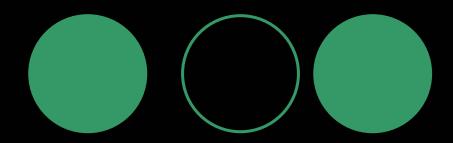




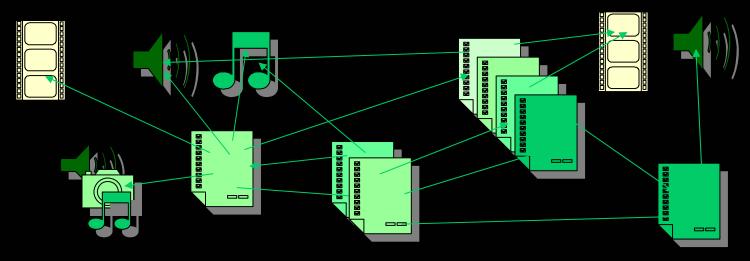
- Grundbegriffe
  - das WWW ist ein "verteiltes Hypermedia-System", das den interaktiven Zugriff unterstützt
  - O Hypermedia ist eine direkte Erweiterung des traditionellen Hypertext:
    - in traditionellen, **nicht Hypertext-basierten** Systemen werden Informationen rein sequentiell, also eindimensional dargestellt



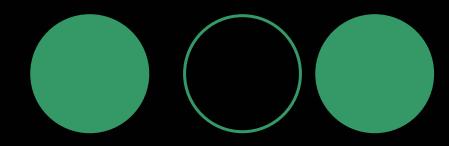




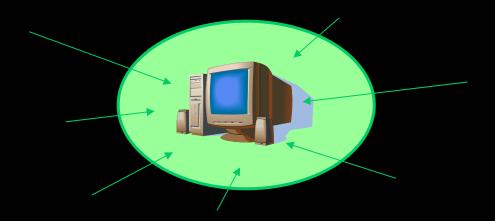
- Grundbegriffe
  - O **Hypermedia** ist eine direkte Erweiterung des traditionellen **Hypertext**:
    - in einem Hypertext basierten System können Dokumente mit anderen Dokumenten/Informationen über so genannte Links miteinander in Verbindung gebracht werden
    - Informationen werden so als Netzwerk von Informationen gespeichert



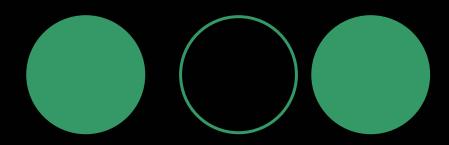




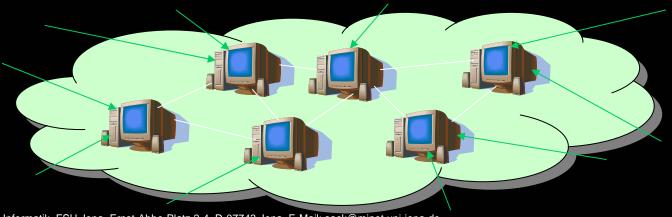
- Grundbegriffe
  - O verteilte und nicht verteilte Hypermedia-Systeme:
    - Nicht verteiltes Hypermedia-System
      - alle Dokumente residieren zentral auf einem einzelnen Rechner
      - Dokumente sind stets lokal verfügbar
      - Links können leicht auf Konsistenz überprüft werden

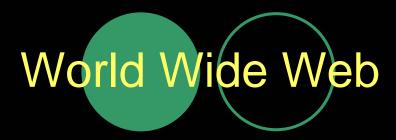


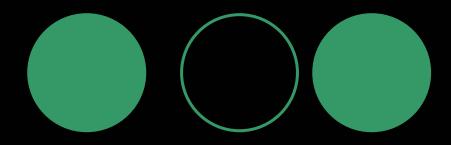




- Grundbegrife
  - O verteilte und nicht verteilte Hypermedia-Systeme:
    - Verteiltes Hypermedia-System
      - Dokumente sind im Web auf (unzähligen) Rechnern verteilt
      - ohne zentrale Kontrolle können
        - neue Dokumente hinzugefügt oder
        - alte umbenannt, geändert oder entfernt werden
      - ohne besondere technische Vorkehrungen sind Links zwischen Dokumenten nicht immer konsistent







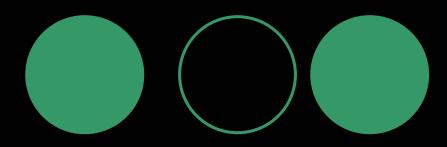
- Grundbegriffe
  - ein Hypermedia-Dokument im WWW heißt "Page"
  - die Hauptseite/Startseite eines Unternehmens oder einer Einzelperson heißt "Homepage"
  - damit Browser als Anzeigeprogramm zwischen Text, Grafiken, Links usw. unterscheiden kann, muss das Format der WWW-Pages sorgfältig definiert und vom Autor stets eingehalten werden

im Web wird als Standarddarstellungsformat HyperText Markup

Language (HTML) benutzt.







- Grundbegriffe
  - damit WWW-Pages eindeutig identifiziert und aufgefunden werden können, wird diesen eine eindeutige Kennung (Adresse) zugeordnet
  - Diese Kennung wird als Uniform Resource Locator (URL) bezeichnet

http://www.zeit.de/index.html

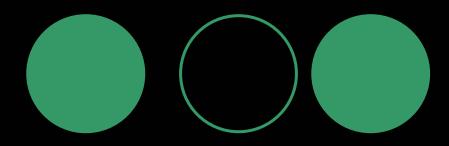




## 3. Internet und WWW (6)

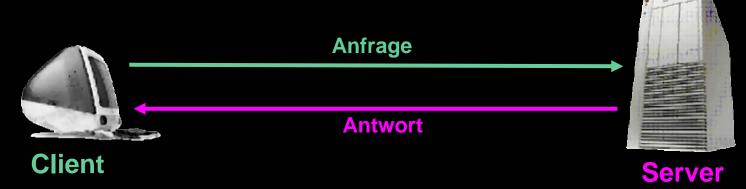
- O World Wide Web
  - Grundbegriffe
  - Browser und WWW-Server
  - eindeutige Identifikation durch URL
  - Kommunikation über HTTP
  - Kodierung mit HTML / CSS
  - O Interaktivität via
    - Dynamisches HTML
    - O cgi-Programme
    - Cookies
  - die Zukunft gehört XML



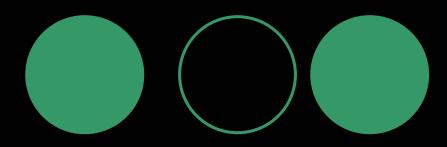


#### Browser und WWW-Server

- Web-Browsing beruht auf dem Client/Server-Prinzip:
- erhält ein Browser einen URL als Eingabe, wird er zum Client
- er kontaktiert dabei den im URL angegebenen Rechner (Server)
- der Browser (Client) fordert den Server auf, ihm das durch den URL spezifizierte Dokument zu senden
- das so erhaltene Dokument wird anschließend vom Browser angezeigt



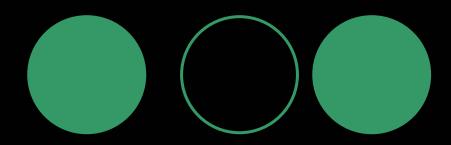




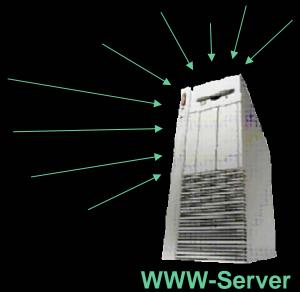
- Browser und WWW-Server
  - Verbindung zwischen Browser und WWW-Server ist stets nur von kurzer Dauer
    - 1. Browser baut eine Verbindung zum Server auf
    - 2. Browser sendet eine Anfrage
    - 3. Server antwortet mit angefordertem Dokument oder Fehlermeldung
    - Browser empfängt angefordertes Dokument und beendet die Verbindung



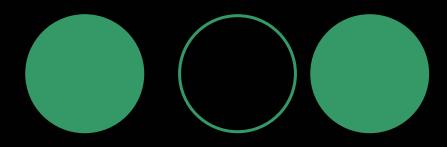




- Browser und WWW-Server
  - WWW-Server sind weniger komplex als Browser
    - Server führen wiederholt einfache Aufgaben aus:
      - Warten auf Eröffnung einer Verbindung durch Browser und Anforderung eines Dokuments,
      - Ausgabe dieses Dokuments bzw. Fehlermeldung und
      - Schließen der Verbindung.
      - erneutes Warten auf Eröffnung
      - •



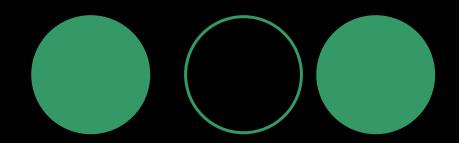




- Browser und WWW-Server
  - O **Browser** haben mehr zu leisten als WWW-Server:
    - Eröffnung der Verbindung zum WWW-Server
    - Anforderung eines Dokuments
    - Einlesen des angeforderten Dokuments
    - Anzeigen des empfangenen Dokuments
    - Reagieren auf Aktionen des Benutzers bei der Bedienung der grafischen Benutzeroberfläche

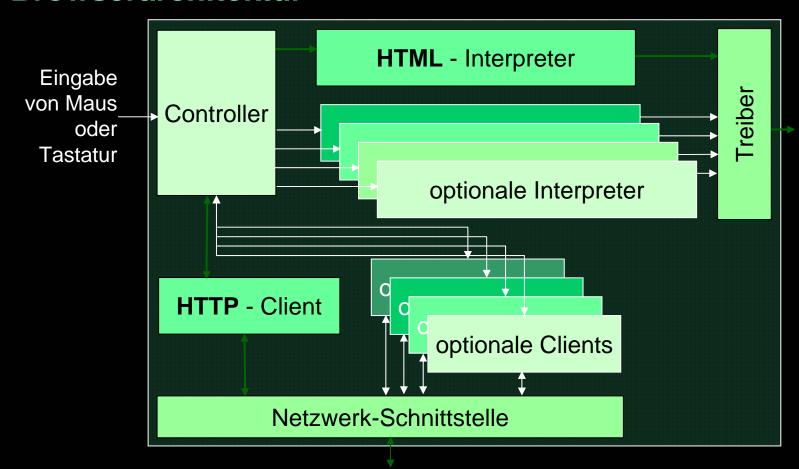


# World Wide Web

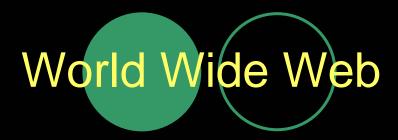


Bildschirm

#### Browserarchitektur



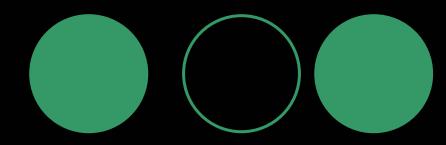
Kommunikation mit entfernten Servern





- Browser und Web-Server
  - Browserarchitektur
    - konzeptionell bestehen Browser aus
      - mehreren Clients,
      - mehreren Interpretern und
      - einem Controller zur Verwaltung.
    - Controller ist zentraler Mittelpunkt des Browsers:
      - **interpretiert** Mausklicks und Tastatureingaben und **aktiviert** andere Komponenten zur Durchführung der vom Benutzer angeforderten Operationen
    - jeder Browser benötigt einen HTML-Interpreter zur formatierten Anzeige von HTML-Dokumenten am Bildschirm.





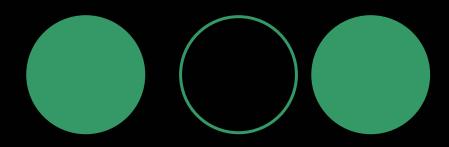
- Browser und WWW-Server
  - Browserarchitektur optionale Clients
    - neben HTTP-Client und HTML-Interpreter kann ein Browser zusätzliche Komponenten zur Durchführung weiterer Aufgaben beinhalten,
      - z.B. FTP-Client für Zugriff auf den Dateitransferdienst
      - E-Mail-Client für Zugriff auf elektronische Mail
      - IRC-Client zur Teilnahme am Chat, usw.
    - Browser versteckt die Details der Benutzung anderer Clients und
    - erlaubt eine automatische Aktivierung des erforderlichen optionalen Diensts
    - Auswahl des entsprechenden Dienstes wird über die Protokollinformation im URL (z.B. http, ftp, ...) gesteuert



## 3. Internet und WWW (6)

- O World Wide Web
  - Grundbegriffe
  - Browser und WWW-Server
  - eindeutige Identifikation durch URL
  - Kommunikation über HTTP
  - Kodierung mit HTML / CSS
  - O Interaktivität via
    - Dynamisches HTML
    - O cgi-Programme
    - Cookies
  - die Zukunft gehört XML





- WWW-Page muss eindeutig identifizierbar sein (Speicherort, Name, Darstellungsart, evtl. integrierte Anwendung)
- in einem verteilten Hypermedia-System recht schwieriges Problem!
- Realisierung im WWW:
  - Kodierung notwendiger Informationen über eindeutige Zeichenkette

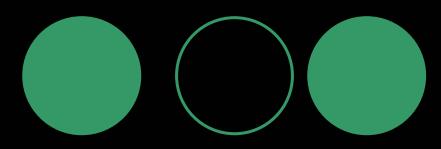
#### **Uniform Resource Identifier – URI**

Ein URI besteht entweder aus einem

Name URN (Uniform Resource Name) oder einem

Adresse URL (Uniform Resource Locator)





Protokoll://computer\_name:port/document\_name

Protokoll

gibt an, wie das Dokument übertragen werden soll (bei HTML → HTTP-Protokoll)

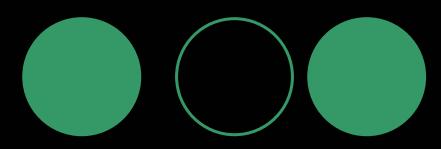
- = Angabe des benutzten Benennungsschemas, wie z.B. des Protokolls.
  - Die IANA (Internet Assigned Numbers Authority) führt eine Liste der "erlaubten" Protokolle und hält die Referenzen auf deren Definitionen

Bsp.: http

ftp

telnet





Protokoll://computer\_name:port/document\_name

Computer-Name identifiziert eindeutig den Rechner, der die

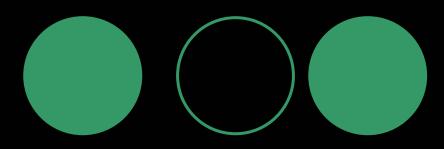
angeforderten Ressourcen bereithält

kann angegeben werden als
 IP-Adresse

**DNS-Name** 

Bsp.: <a href="www.uni-jena.de">www.uni-jena.de</a>
<a href="mailto:189.139.99.1">189.139.99.1</a>
<a href="mailto:ipcx23.ibm.com">ipcx23.ibm.com</a>





Protokoll://computer\_name:port/document\_name

**Port-Name** 

identifiziert den Port am Server-Rechner, der mit dem jeweils angeforderten Service verbunden ist

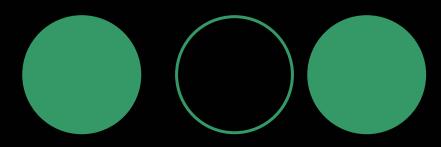
 Standard HTTP-Service liegt stets auf Port 80 und muss daher nicht mit angegeben werden

Bsp.: www.uni-jena.de:80 189.139.99.1:6767

Bemerkung: auf einem Server-Rechner können parallel mehrere

verschiedene WWW-Server (Software) arbeiten





Protokoll://computer\_name:port/document\_name

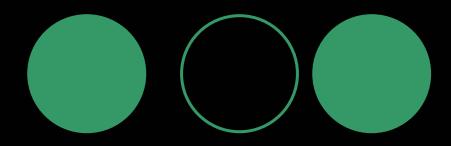
**Document-Name** identifiziert das angeforderte Dokument eindeutig auf dem angegebenen Server-Rechner

 kann z.B. aus einem Verzeichnisnamen mit angefügtem Dateinamen bestehen

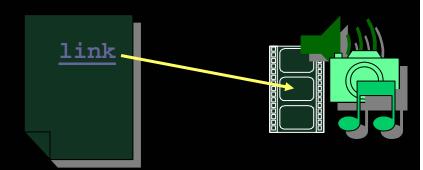
Bsp.: test/ss01/index.html

Bezeichnet die Datei index.html im Verzeichnis test/ss01





- HTML nutzt URLs als Hypermedia-Referenz auf Objekte
  - referenziertes Objekt muss sich nicht lokal auf dem Rechner befinden
- referenzierbare Objekte:
  - andere HTML-Dokumente
  - Grafik
  - beliebige Dokument-Typen,
  - ausführbare Programme,
  - •



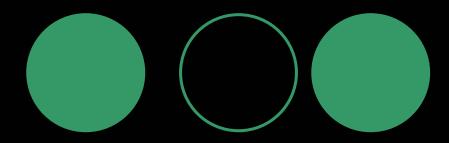
- Objekttyp wird durch Dateityp festgelegt
  - z.B: bild.jpg



## 3. Internet und WWW (6)

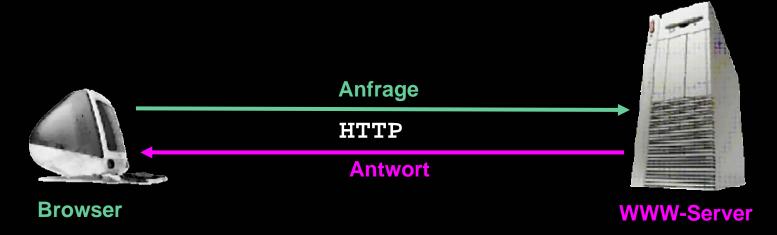
- O World Wide Web
  - Grundbegriffe
  - Browser und WWW-Server
  - eindeutige Identifikation durch URL
  - Kommunikation über HTTP
  - Kodierung mit HTML / CSS
  - O Interaktivität via
    - Dynamisches HTML
    - O cgi-Programme
    - Cookies
  - die Zukunft gehört XML



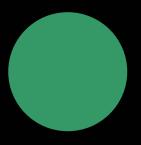


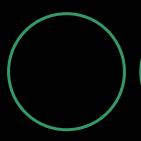
### HyperText Transport Protocol

- Die Interaktion zwischen Browser und WWW-Server erfolgt über das HyperText Transfer Protokoll - HTTP
- HTTP gibt das genaue Format
  - für eine **Anfrage** vom Browser an einen Server und
  - für eine Antwort eines Servers an den Browser vor











## HyperText Transport Protocol

HTTP-Operationen (Methoden)



Browser

**GET** Anforderung einer Ressource

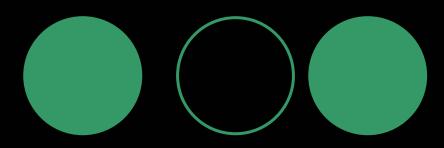
vom Server

POST/PUT

Übermittlung von Informationen

vom Browser zum Server





**WWW-Server** 

## HyperText Transport Protocol

Typische HTTP-Kommunikation

GET http://www.uni-jena.de/index.html HTTP/1.0



**Browser** 

HTTP/1.1 200 Document follows

Date: Tue, 10-May-2004 12:00:01 GMT

Content Type: text/html

Last Modified: Mon, 09-Sep-2002 19:33:00 GMT

Content Length: 8164

.....angefordertes document....







- HyperText Transfer Protocol
  - O HTTP Status-Codes
    - HTTP-Statuscodes dienen zur Kommunikation von Status- oder Fehlermeldungen zwischen Browser und WWW-Server
    - Einteilung in Gruppen:

1xx - Informational	vorläufige Antwort, Server hat Anfrage erhalten und bearbeitet diese gerade
2xx - Successful	Anfrage wurde vom Server empfangen, ver- standen und akzeptiert
3xx – Redirection	Anfrage konnte nicht (vollständig) vom Server bearbeitet werden, Verweis auf anderen Server
4xx – Client Error	Anfrage konnte nicht bearbeitet werden, da Fehler auf Client-Seite (z.B. falscher URL)
5xx – Server Error	Anfrage konnte nicht bearbeitet werden, da Fehler auf Server-Seite



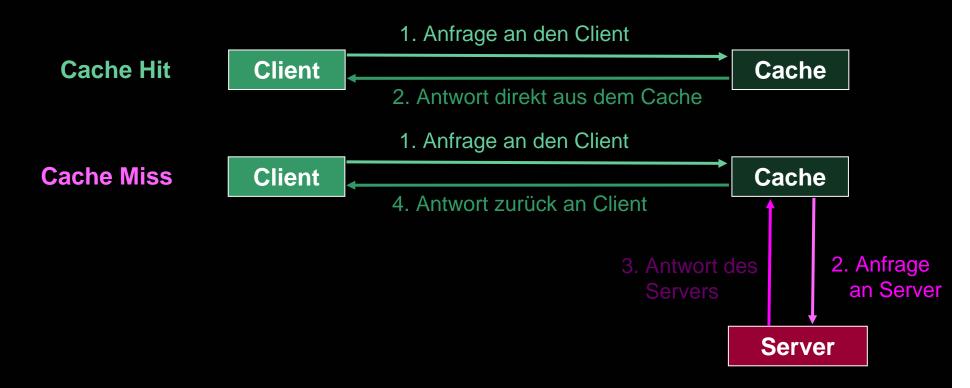


- HyperText Transfer Protocol
  - O HTTP Caching
    - Um unnötige Datenübertragungen zu vermeiden, verfügt der Browser über einen intelligenten Zwischenspeicher (Cache) für bereits empfangene Dokumente
    - Ist ein anzuforderndes Dokument bereits im Cache des Browsers vorhanden, kann dieses direkt aus dem Cache bezogen werden, ohne den Server zu kontaktieren
    - Cache-Speicher können unterschiedlich realisiert werden:
      - auf Client-Seite
      - eigenständig
      - auf Server-Seite
      - mit unterschiedlichen Caching-Strategien

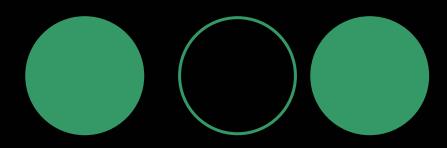




- HyperText Transfer Protocol
  - HTTP Caching Cache Operationen





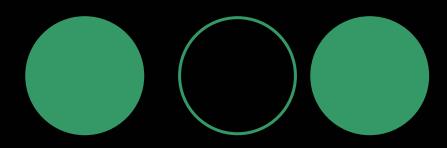


- HyperText Transfer Protocol
  - Clientseitiger Cache



- Cache liegt beim Client
- Speichert jedes übertragene Dokument
- History-Funktion (Zurück) des Browsers





- HyperText Transfer Protocol
  - Eigenständiger Cache



- Cache ist zwischen Client und Server platziert
- kann mit Client gemeinsam denselben Rechner nutzen, ist aber logisch eigenständig
- muss am Browser speziell konfiguriert werden
- Cache-Hierarchie möglich



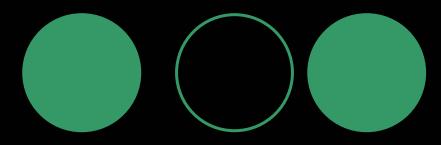


- HyperText Transfer Protocol
  - Serverseitiger Cache

Client Cache Server

- Cache ist direkt beim Server platziert
- speichert Antworten, die der Server erst auf Anfrage hin aufwändig berechnen muss bzw. die von Clients besonders häufig angefragt werden





- HyperText Transfer Protocol
  - Caching-Strategien
    - Wann soll eine angeforderte Ressource direkt vom WWW-Server angefordert werden und wann soll diese Anfrage aus dem Cache beantwortet werden?
    - HTTP sieht dazu bestimmte Datenfelder im Header der HTTP-Nachricht vor.

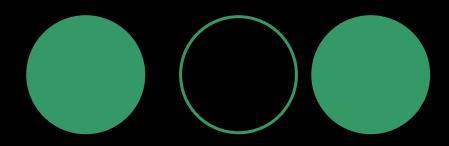
**Last Modified** wann wurde die Ressource zuletzt verändert

**Expires** wie lange ist die Ressource gültig?

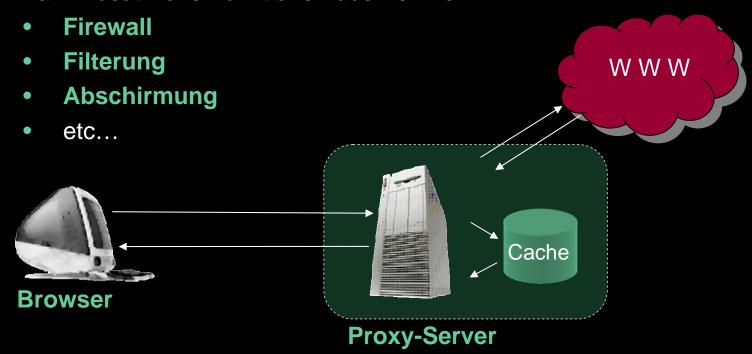
ab wann muss diese vom Server direkt

nachgeladen werden?





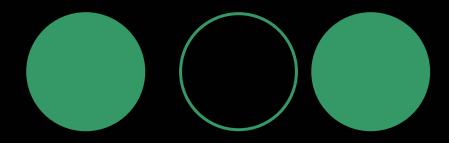
- HyperText TransferProtocol
  - O Proxy-Server
    - Als Cache wird oft ein so genannter Proxy-Server eingerichtet
    - Kann zusätzliche Funktionen übernehmen





- O World Wide Web
  - Grundbegriffe
  - Browser und WWW-Server
  - eindeutige Identifikation durch URL
  - Kommunikation über HTTP
  - Kodierung mit HTML / CSS
  - Interaktivität via
    - Dynamisches HTML
    - O cgi-Programme
    - Cookies
  - die Zukunft gehört XML

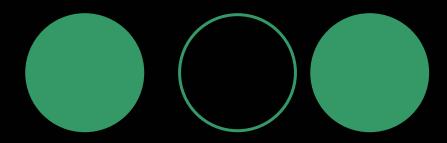




- HTML und CSS
  - HyperText Markup Language
    - Hypermedia-Dokumente werden im WWW in der HyperText Markup Language (HTML) kodiert

```
🚂 Source of: file:///C:/Dokumente%20und%20Einstellungen/admin/Eigene%20Dateien/download/Heise%20News-Ticker%20%2075%20Jah...
File Edit View Help
<html><head>
                                                                                                    or or leases
<meta name="ID" content="38236">
<meta name="Author" content="Andreas Stiller">
<meta name="datum" content="20030703">
<meta name="objekt" content="ct">
<meta name="titel" content="75 Jahre Farbfernsehen">
                                                                                                     CREWIATY EIRES, LINE
ACTION EIRES
<meta name="description" content="Heute vor 75 Jahren führte der schottische Erfinder
John Logie Baird (1888-1946) zum ersten mal eine Farbfernsehübertragung vor.">
<meta name="Keywords" content="75 Jahre Farbfernsehen">
<meta name="copyright" content="@ Heise Zeitschriften Verlag">
<link href="http://www.heise.de/favicon.ico" rel="shortcut icon"><title>Heise News-Ticker: 7
                                                                                                     trockener Tuckers
<!-- base --></head>
<body leftmargin="10" topmargin="2" marginwidth="10" marginheight="2" text="#000000" bgcolor</pre>
<div align="center">
                                                                                                     MEASUREMENT BY
                                                                                                s bearings.
<a href="http://www.heise.de/"><img src="Heise%20News-Ticker%20%2075%20Jahre%20Farbf
<img src="Heise%20News-Ticker%20%2075%20Jahre%20Farbfernsehen files/hardware.gif" width="1"</pre>
```





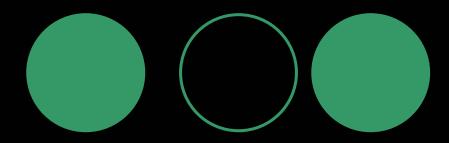
### HTML und CSS

- "Markup" bedeutet Auszeichnen und meint damit das Kennzeichnen von bestimmten Abschnitten mit gewissen Attributen oder Anweisungen
- will man einen Textabschnitt fett darstellen, müssen Anfang und Ende des Abschnittes mit dem "Fett-Markup" gekennzeichnet werden

die einzelnen Markup-Elemente heißen "Tags"

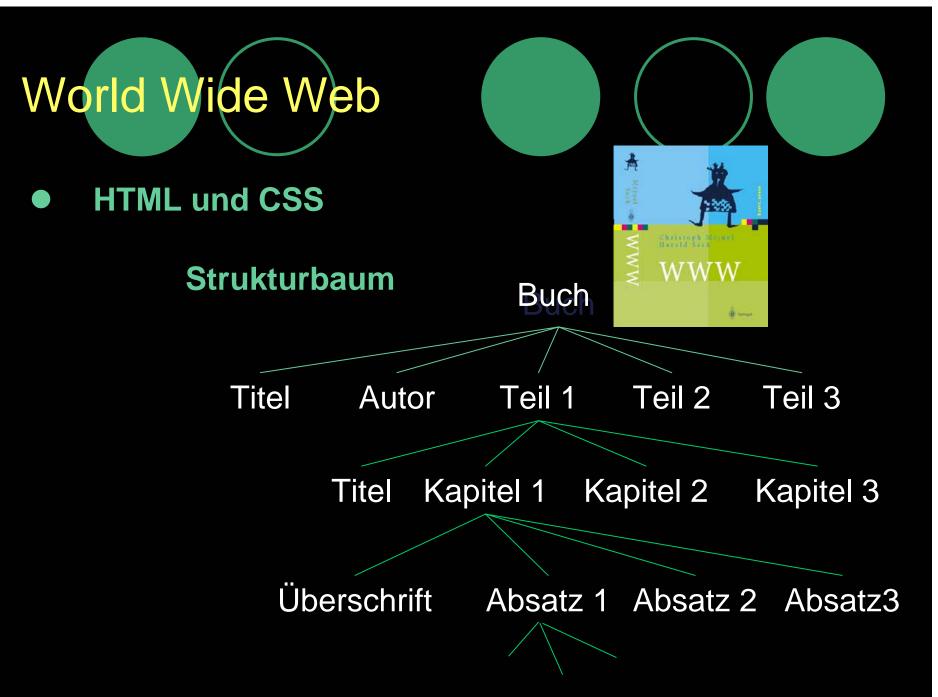




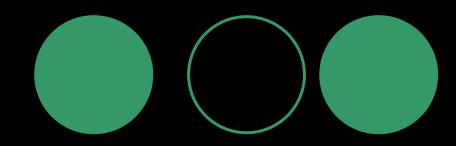


- HTML und CSS
- Beispiel:
  - O <b> Dies ist ein fettgedruckter Text </b>
    - <b> markiert hierbei den Anfang und
    - </b> das Ende des Fettdrucks
- normalerweise enthält Markup keine ausführlichen Formatierungsanweisungen
- es werden nur allgemeine Richtlinien zur Strukturierung der Dokumente beschrieben
- Einzelheiten der Formatierung werden dem Browser überlassen



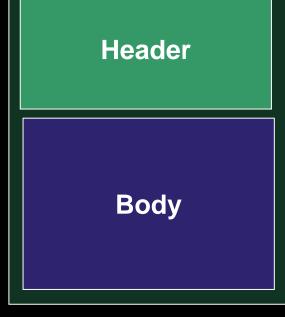






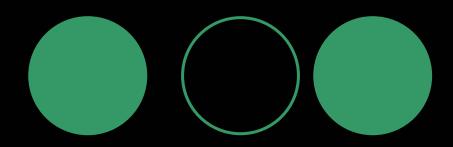
### HTML und CSS

- o HTML-Dokument besteht aus:
  - o dem Kopf "Header" enthält Informationen über das Dokument, z.B. Titel, oder Stichwörter über den Dokumenteninhalt usw. und
  - o dem Rumpf "Body" enthält eigentliche Informationen, die der Autor hier über Markups/Tags mit einer Struktur und Formathinweisen versehen kann
- o HTML erlaubt z.B. die Auszeichnung von
  - o Überschriften,
  - o Listen,
  - o Grafiken,
  - o Links ...

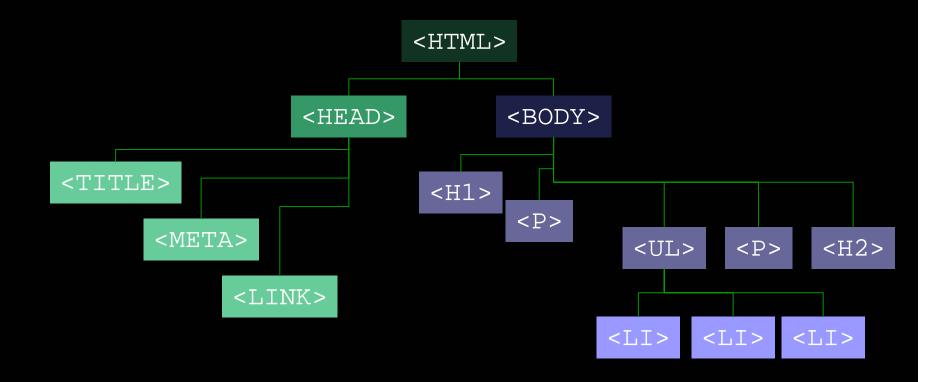


**HTML-Dokument** 

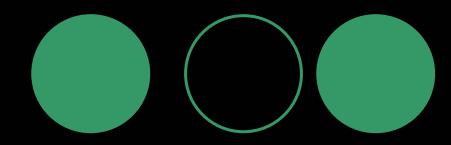




- HTML und CSS
  - Aufbau eines HTML-Dokuments





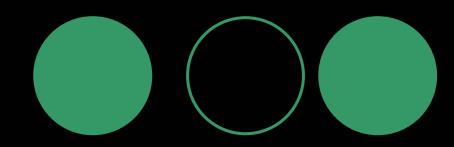


- HTML und CSS
  - O HTML Hyperlinks
    - Hypermedia-Referenzen werden im Browser als auswählbares (anklickbares) HTML-Element dargestellt
    - **jedes HTML-Element** (Wort, Satz, Absatz, Bild) kann als Hypertext-Referenz verwendet werden
    - das HTML-Element wird dazu zusammen mit dem URL des verlinkten Dokuments von zwei "Ankern" (Anchor) - den Tags <a>a> und </a> - eingerahmt

<a href="http://www.uni-jena.de/~sack/index.html">
 Dies ist ein Link
</a>

Dies ist ein Link

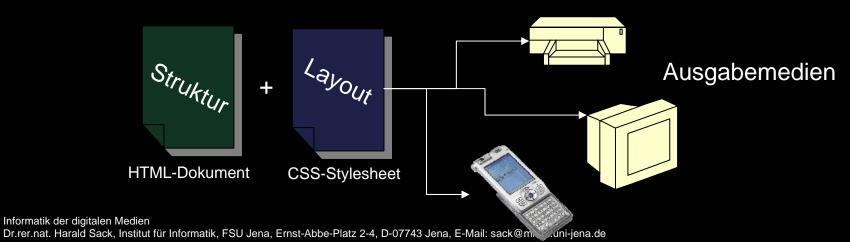




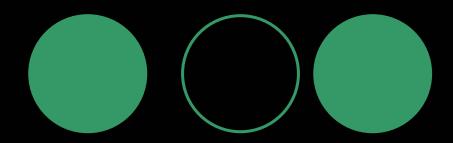
- HTML und CSS
  - HTML und Formatierung mit CSS
    - HTML legt nur die Dokumentenstruktur fest (idealerweise)
    - Zur Formatierung der Darstellung der einzelnen HTML-Strukturelemente (Überschriften, Listen, Tabellen, etc.) dient eine spezielle Formatierungssprache:

#### **Cascading Style Sheets (CSS)**

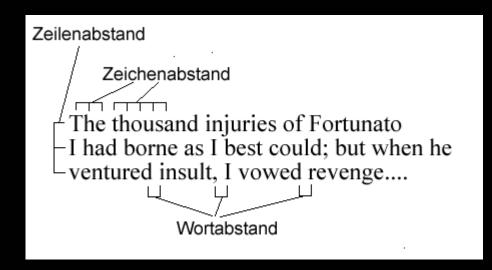
 Das CSS legt das Layout eines HTML-Dokuments für ein bestimmtes Ausgabemedium (Bildschirm, Drucker, etc.) fest.



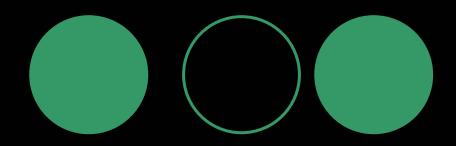




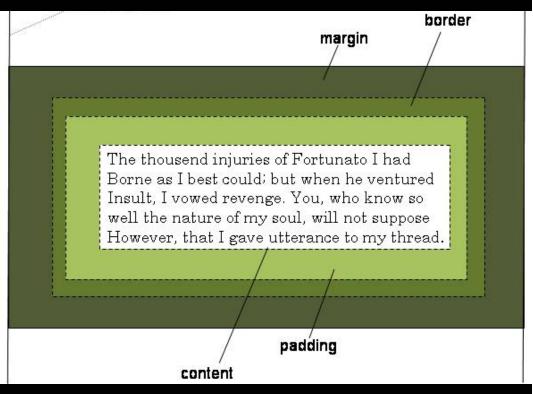
- HTML und CSS
  - Formatierung mit CSS
    - CSS-Stylesheets bestimmen
      - Schriftattribute
        - fett, unterstrichen, kursiv, etc.
        - Zeichen-/Wortabstand
        - Zeilenabstand





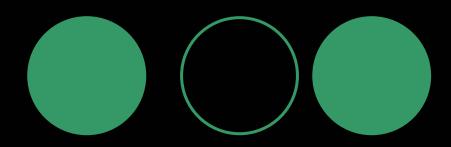


- HTML und CSS
  - Formatierung mit CSS
  - CSS-Stylesheets bestimmen
    - Bemaßungen
      - Ränder
      - Zwischenraum
      - Position

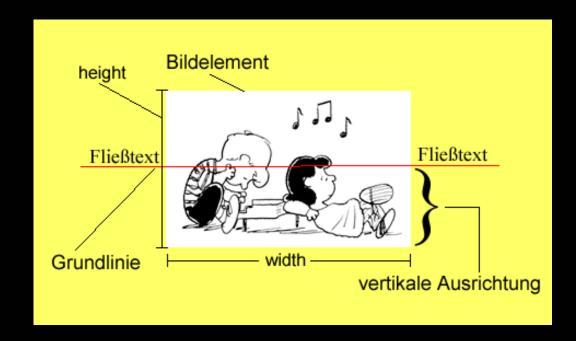


Informatik der digitalen Medien



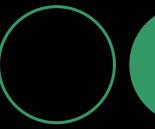


- HTML und CSS
  - Formatierung mit CSS
  - CSS-Stylesheets bestimmen
    - Objektausrichtung zum Text



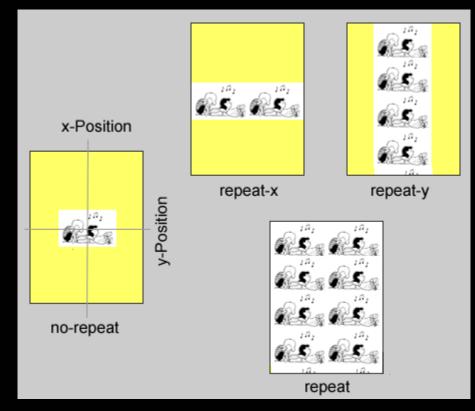




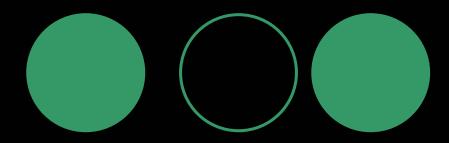




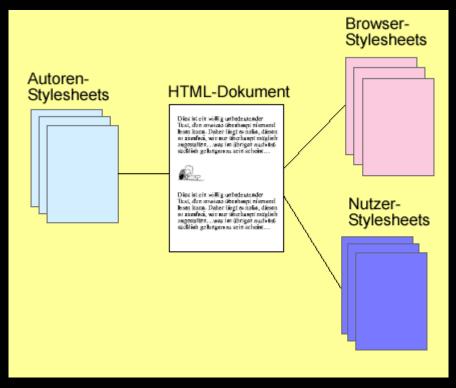
- HTML und CSS
  - Formatierung mit CSS
  - CSS-Stylesheets bestimmen
    - Hintergrundgestaltung







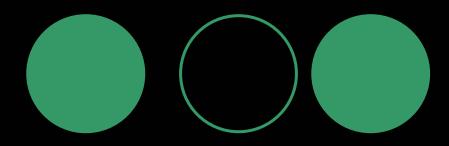
- HTML und CSS
  - Formatierung mit CSS
    - Für ein HTML-Dokument können verschiedene, konkurrierende CSS-Stylesheets angegeben werden
    - unterscheide
      - Autorenstylesheet
      - Browserstylesheet
      - Nutzerstylesheet





- O World Wide Web
  - Grundbegriffe
  - Browser und WWW-Server
  - eindeutige Identifikation durch URL
  - Kommunikation über HTTP
  - Kodierung mit HTML / CSS
  - O Interaktivität via
    - Dynamisches HTML
    - O cgi-Programme
    - Cookies
  - die Zukunft gehört XML

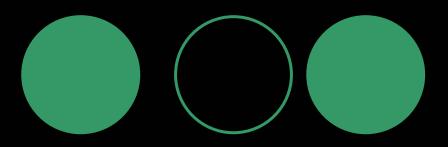




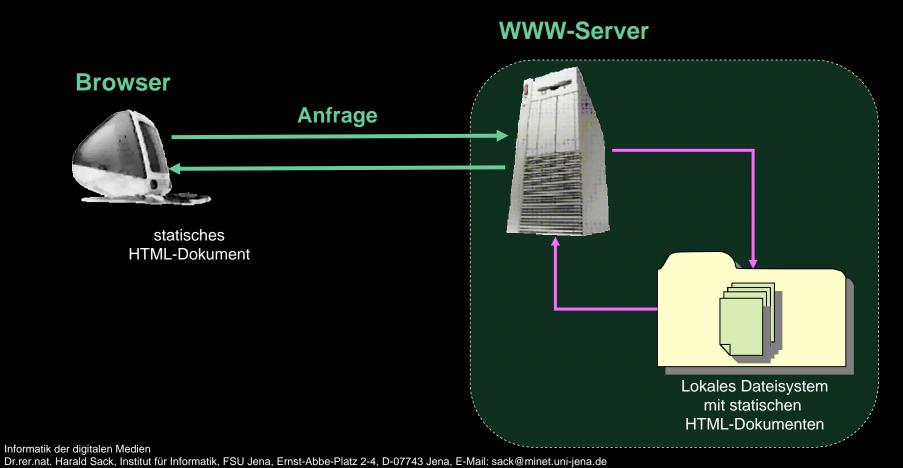
### Interaktivität - Dynamisches HTML

- HTML-Dokumente können statisch,
  - d.h. bereits vorgefertigt auf dem WWW-Server vorliegen oder
- dynamisch erzeugt werden,
  - d.h. das vom Browser angefragte HTML-Dokument wird erst auf die Anfrage hin erzeugt.
  - ist sinnvoll, wenn z.B. Anfragen aus einer Datenbank beantwortet werden müssen
    - z.B. Warenkataloge (vgl. amazon.de)
    - z.B. Suchmaschinen (vgl. google.de)
    - z.B. Zeitungen (vgl. heise.de)

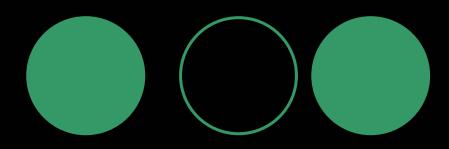




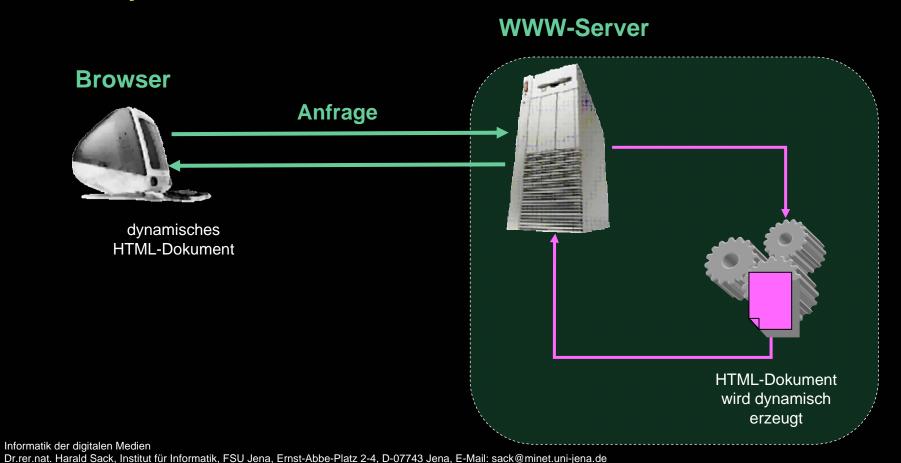
- Interaktivität Dynamisches HTML
  - Statisches HTML



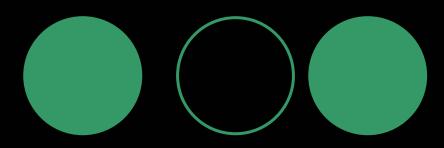




- Interaktivität dynamisches HTML
  - O Dynamisches HTML

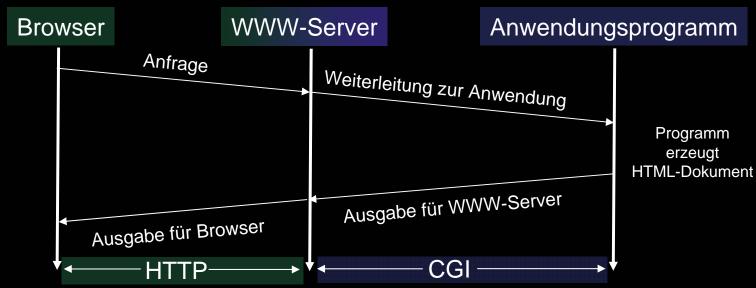




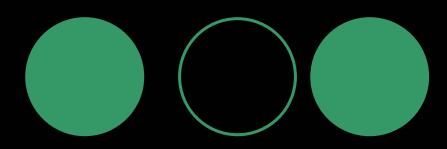


- Interaktivität- CGI Schnittstelle
  - WWW-Server bietet Standard-Schnittstelle für serverseitige Programme, mit deren Hilfe das dynamische HTML-Dokument erzeugt wird

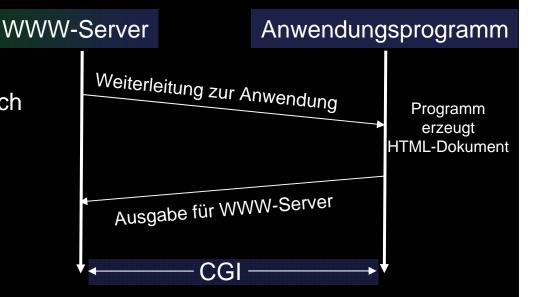
Common Gateway Interface (CGI)





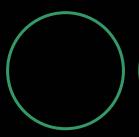


- Interaktivität- CGI Schnittstelle
  - WWW-Server und Anwendung kommunizieren über cgi-Schnittstelle
  - WWW-Server kann auch Daten von Browser entgegennehmen und an die Anwendung weiterleiten











### Interaktivität - Cookies

- Wie erkennt mich der WWW-Server eigentlich wieder?
- HTTP ist eigentlich ein zustandsloses Protokoll, d.h.
  - über früher stattgefundene Transaktionen hat HTTP keine Information
- Informationen über früher stattgefundene Transaktionen sind aber wichtig z.B. für Shopping-Portale (vgl. Warenkorb)
- Diese Information muss daher bei Client und Anwendungsprogramm abgespeichert werden
- Die beim Client gespeicherte Information wird als Cookie bezeichnet







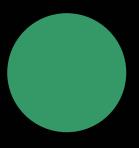




### Interaktivität - Cookies

- Was kann man mit Cookies alles machen?
  - Identitätsüberprüfung des Nutzers
  - Feststellen, ob der Nutzer den WWW-Server bereits schon einmal besucht hat
  - Erstellen von Kunden- und Nutzer-Profilen
- Was genau ist ein Cookie?
  - Zeichenkette, die zwischen Client und WWW-Server kommuniziert wird
  - Zum Austausch von Informationen, die nicht durch das HTTP-Protokoll ausgetauscht werden können
- Sind Cookies ein Sicherheitsrisiko?





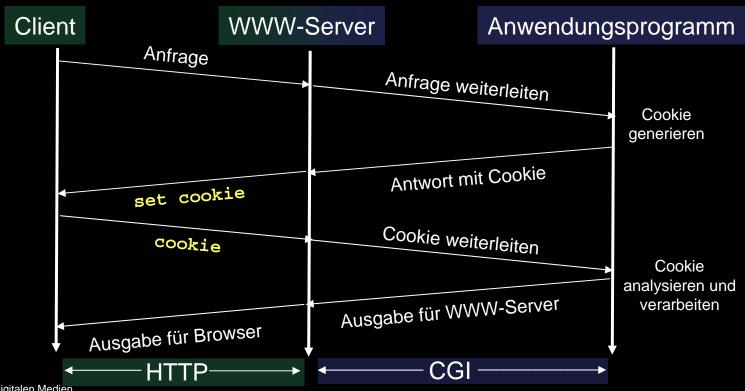




Interaktivität - Cookies

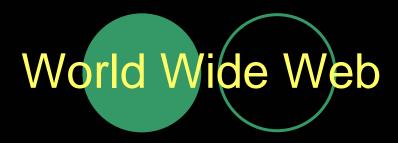


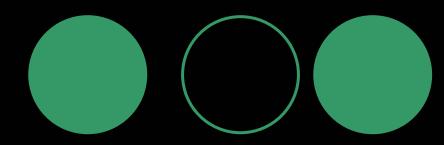
Wie funktionieren Cookies?



Informatik der digitalen Medien

Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-07743 Jena, E-Mail: sack@minet.uni-jena.de





- Interaktivität Cookies
  - Sind Cookies ein Sicherheitsrisiko?
    - Client darf nur Cookies an WWW-Server weitergeben, die auch von diesem stammten
    - fremde Cookies dürfen niemals weitergegeben werden
    - Aber:
      - Alle Cookies werden auf Client-Seite in derselben Datei abgespeichert → daher ist ein Missbrauch denkbar
    - Daher:
      - Sorgsamer Umgang mit der Cookie-Datei ist empfohlen!



- O World Wide Web
  - Grundbegriffe
  - Browser und WWW-Server
  - eindeutige Identifikation durch URL
  - Kommunikation über HTTP
  - Kodierung mit HTML / CSS
  - O Interaktivität via
    - Dynamisches HTML
    - O cgi-Programme
    - Cookies
  - die Zukunft gehört XML





- XML Extensible Markup Language
  - Warum HTML nicht ausreicht
    - Kann ein Programm automatisch ein WWW-Dokument verstehen?

Max Mustermann Sesamstr. 49a 93123 Bad Sulzdetfurth Ein Mensch identifiziert diese Zeilen sofort als Adressangabe

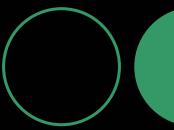
egal, ob sie von anderem Text umgeben sind oder nicht

Aber für ein Programm handelt es sich erst einmal nur um einen beliebigen Text

Was kann man tun, damit ein Programm ein WWW-Dokument Besser versteht?









- XML Extensible Markup Language
  - Warum HTML nicht ausreicht
    - Was kann man tun, damit ein Programm ein WWW-Dokument besser versteht?

Max Mustermann Sesamstr. 49a 93123 Bad Sulzdetfurth

Adressen müssten als solche **gekennzeichnet** werden

z.B.

<Adresse>

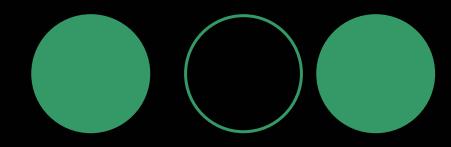
Max Mustermann Sesamstr. 49a 93123 Bad Sulzdetfurth

</Adresse>

#### **Problem:**

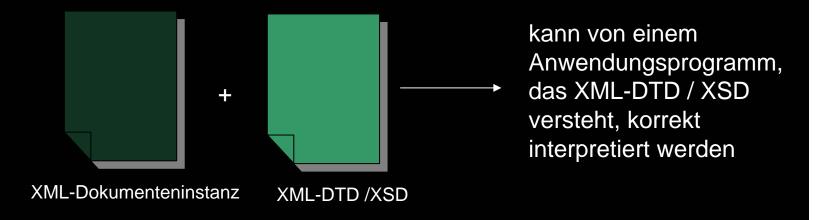
HTML lässt die Definition neuer Tags nicht zu....

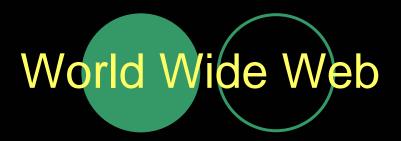


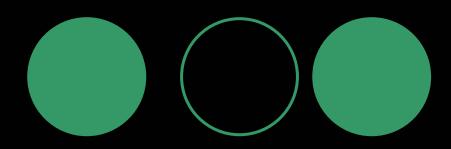


### XML – Extensible Markup Language

- Was XML alles kann
  - XML gestattet die Definition beliebiger neuer Tags
  - die Definition neuer Tags erfolgt in einem speziellen Dokument,
    - der Document Type Definition (DTD) oder
    - einer XMLSchema Definition (XSD)







- XML Extensible Markup Language
  - Was XML alles kann

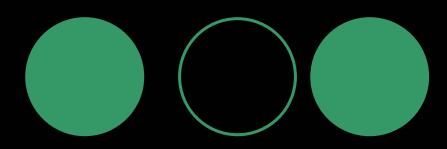
Max Mustermann Sesamstr. 49a 93123 Bad Sulzdetfurth

```
<Adresse>
  <Vorname> Max </Vorname>
  <Nachname> Mustermann </Nachname>
  <Straße> Sesamstr. </Straße>
  <Hausnummer> 49a </Hausnummer>
  <PLZ> 93123 </PLZ>
  <Ort> Bad Sulzdetfurth </Ort>
  </Adresse>
```

#### **Wohlgeformte XML-Dokumente:**

Syntaktisch korrekt, d.h. alle angefangenen Tags werden wieder geschlossen und sind stets korrekt geschachtelt.





- XML Extensible Markup Language
  - Was XML alles kann

```
<Adresse>
  <Vorname> Max </Vorname>
  <Nachname> Mustermann </Nachname>
  <Straße> Sesamstr. </Straße>
  <Hausnummer> 49a </Hausnummer>
  <PLZ> 93123 </PLZ>
  <Ort> Bad Sulzdetfurth </Ort>
  </Adresse>
```

```
<!ELEMENT Adresse (Vorname,
   Nachname, Straße, Hausnummer,
   PLZ, Ort)>

<!ELEMENT Vorname CDATA>
<!ELEMENT Nachname CDATA>
<!ELEMENT Straße CDATA>
....
```

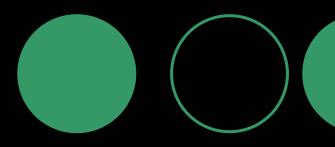
**XML-DTD** 

**XML** 

#### validierte XML-Dokumente:

wohlgeformt und den Definitionen der zugeordneten DTD entsprechend





- XML Extensible Markup Language
  - Das Semantische Web WWW 2.0
    - Aber um den Inhalt eines Dokuments wirklich zu verstehen ist mehr nötig!
      - Kontextwissen
      - Regeln
      - Weltwissen

...Wenn Max Mustermann der Name eines real existierenden Menschen ist, dann hat dieser mit großer Wahrscheinlichkeit mindestens einen Wohnsitz, der durch die Angabe von Straße, Hausnummer, PLZ und Wohnort hinreichend spezifiziert werden muss...



das WWW der nächsten Generation soll dieses Wissen nutzen



- XML Extensible Markup Language
  - Das Semantische Web WWW 2.0



- WWW-Dokumente des Semantic Web sollen von autonom agierenden Agenten (Anwendungsprogrammen) automatisch "verstanden" werden
- Agenten sollen in der Lage sein, autonom Entscheidungen im Sinne des Nutzers zu treffen, um gezielt und intelligent Aufträge des Nutzers ausführen zu können

z.B. eine Suchmaschine, die tatsächlich das findet, was ich suche.....



- O World Wide Web
  - Grundbegriffe
  - Browser und WWW-Server
  - eindeutige Identifikation durch URL
  - Kommunikation über HTTP
  - Kodierung mit HTML / CSS
  - O Interaktivität via
    - Dynamisches HTML
    - O cgi-Programme
    - Cookies
  - die Zukunft gehört XML



#### Literatur



Ch. Meinel, H. Sack: *WWW– Kommunikation, Internetworking, Web-Technologien,* Springer, 2004.