

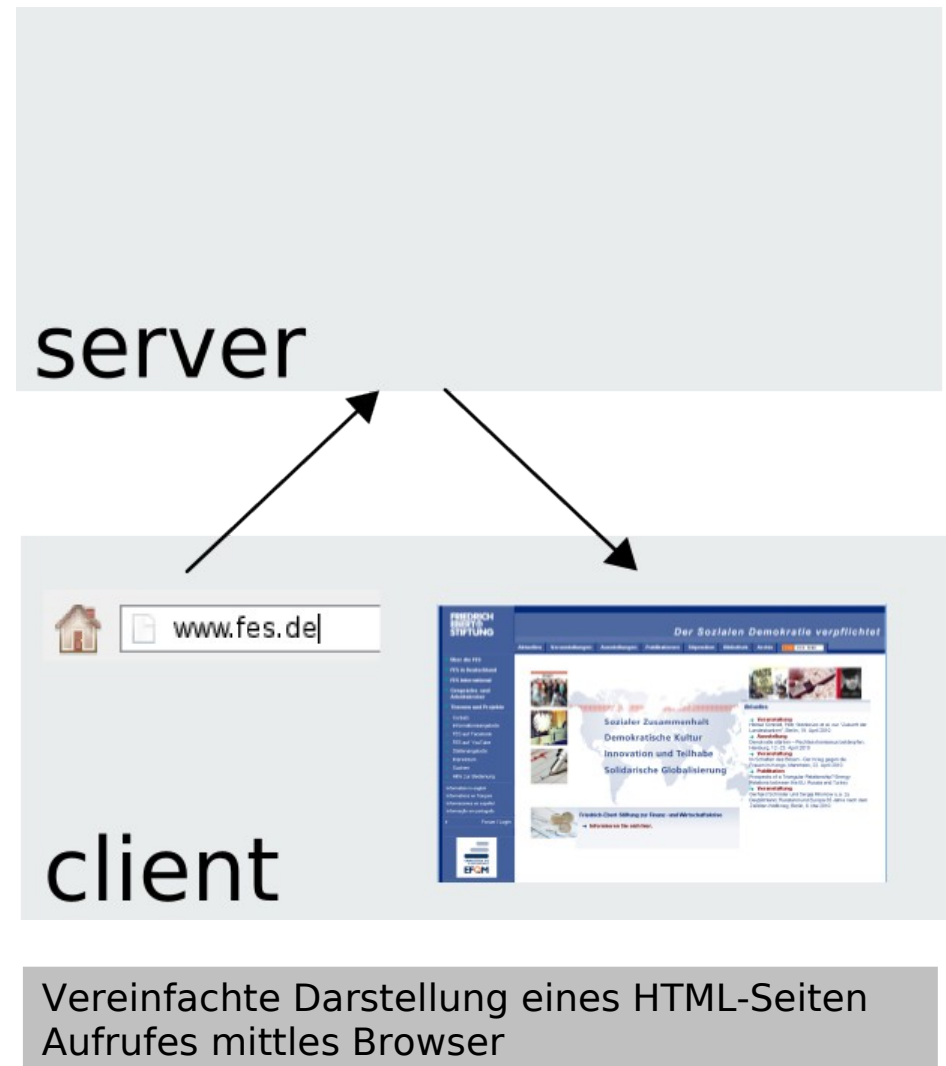
Zwischen Client und Server / Sekundärdaten und die Möglichkeit ihrer Auswertung

Sebastian Vesper, FES

AWV Arbeitskreissitzung 6.2
„Dokumentation und Archivierung von Webpräsenzen“
20. April 2010, Bonn

Internet und Datentransfer

- WWW als Teil des Internets
- Webbrowser zur Nutzung
- HTTP (Hypertext Transport Protocol) dient der Übertragungen von Informationen im World Wide Web
 - Anfrage:
client -> server
Im Browser per URL (Uniform Resource Locator)
 - Antwort:
server -> client
übermittelt HTML-Seite



HTTP

- HTTP = Hypertext Transfer Protocol
- Protokoll: Spezifikation der Kommunikation
 - „Ein Netzwerkprotokoll (auch Netzprotokoll, Übertragungsprotokoll) ist eine exakte Vereinbarung (Protokoll), nach der Daten zwischen Computern bzw. Prozessen ausgetauscht werden, die durch ein Netz miteinander verbunden sind (verteiltes System). Die Vereinbarung besteht aus einem Satz von Regeln und Formaten (Syntax), die das Kommunikationsverhalten der kommunizierenden Instanzen in den Computern bestimmen (Semantik).“
(Seite „Netzwerkprotokoll“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 21. Februar 2010, 13:00 UTC. URL: <http://de.wikipedia.org/w/index.php?title=Netzwerkprotokoll&oldid=70960342>)
- Entwickelt 1989 von Tim Berners-Lee, Roy Fielding u.a.
- Entwicklung von HTTP wird vom der IETF (Internet Engineering Task Force) und dem W3C (World Wide Web Consortium) koordiniert
- Spezifikationen dokumentiert in RFCs (Requests for Comments)
 - HTTP 1.0: RFC 1945 (<http://tools.ietf.org/html/rfc1945>)
 - HTTP 1.1: RFC 2616 (<http://tools.ietf.org/html/rfc2616>)

HTTP & TCP/IP

- HTTP

- Protokoll auf Anwendungsebene, im OSI-Modell Schicht 7-5 (Open Systems Interconnection Reference Model)
 - Populärste Implementierung auf TCP/IP (Transmission Control Protocol / Internet Protocol)
 - Nachrichtenorientierte Kommunikation (request-response), dadurch z.B. Zustandslosigkeit

Anwendungen (Browser ...)

HTTP

TCP

IP

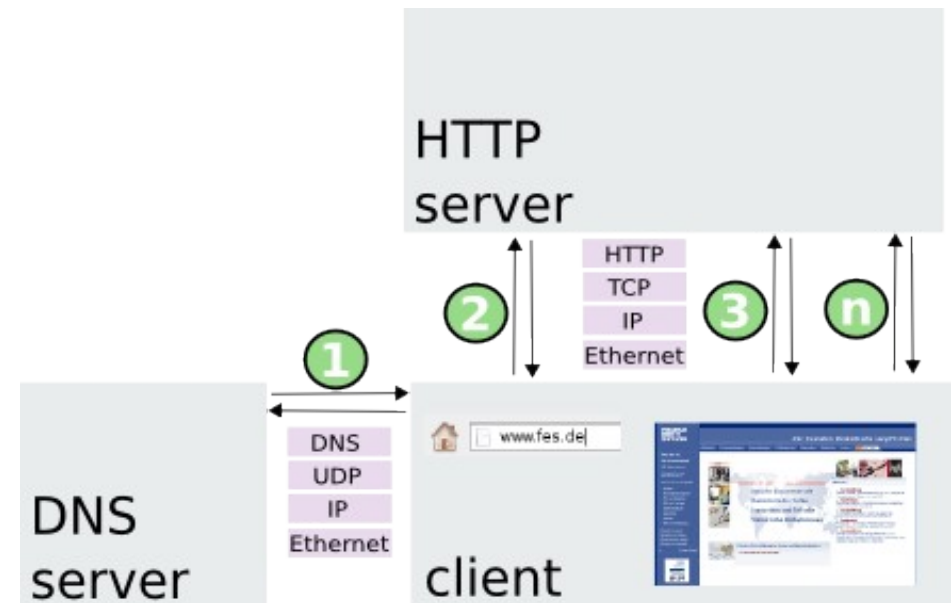
Ethernet

HTTP Nachrichtenstruktur

- HTTP-Nachrichten bestehen aus einem Nachrichtenkopf (Message Header) und einem Nachrichtenkörper (Message Body)
- Nachrichtenkopf beinhaltet Zusatzinformationen (Sekundärdaten)
 - Client-seitig u.a. die eingestellte Sprache (de,en) , die möglichen Transportmethoden (gzip), unterstützte Zeichensätze (UTF-8, ISO-8859-1) und Informationen zu Verbindungsarten
 - Server-seitig werden neben dem Datum, Serversoftware Informationen zum Nachrichtenkörper übermittelt. Die Art „Content-Type“ (Text, Binärdaten), Größe „Content-Length“ (in Byte), Kodierung „Content-Encoding“ usw.
- Nachrichtenkörper beinhaltet die eigentlichen Daten, welche im Browser angezeigt werden (Primärdaten)

Datentransfer im Detail

- Aufruf einer URL im Browser löst folgende Aktionen aus:
 - Lookup der Domain mittels DNS (Domain Name System) per UDP (User Datagram Protocol)
z.B. google.de -> 209.85.229.104
 - HTTP Anfrage senden
"GET / HTTP/1.1"
 - HTTP Antwort parsen
 - Message Header auswerten
 - Message Body auswerten
 - HTML parsen
 - Referenzen auf externe CSS-, JavaScript- oder Binärdateien (Bilder, Flash..) auflösen und diese nachladen



HTTP Eigenschaften

- Request-Methoden u.a.:
 - GET um Daten anzufordern
 - POST um Daten an den Server zu senden
 - HEAD übermittelt nur den HTTP-Header
- HTTP Statuscodes u.a.:
 - 200 OK (Anfrage erfolgreich bearbeitet)
 - 404 Not Found (Ressource nicht gefunden)
 - 304 Not Modified (Ressource nicht modifiziert seit letzten Aufruf)

Visualisierung von HTTP Nachrichten

- Browser-basiert lassen sich HTTP-Nachrichten mittels
 - Firebug für Mozilla Firefox (<http://getfirebug.com>)
oder
 - Google Chrome (<http://www.google.de/chrome>)
analysieren (insbesondere bzgl. des Kopfteiles d. Nachrichten).
- Screenshot Firebug:

