

Wie funktioniert der Internet-Dienst WWW?

Das Medium Internet mit seinem Dienst World Wide Web (WWW) ist genau vor erst 17 Jahren der Weltöffentlichkeit vorgestellt worden. Bis heute hat sich diese Technologie in einem atemberaubenden Tempo weiterentwickelt. Wie funktioniert genau der Internet-Dienst WWW? Dieser Artikel gibt eine fundierte Einführung.

Das World Wide Web und seine Geschichte

Genau gesagt ist WWW ein Dienst des Medium Internet (globales Rechnernetz auf Basis der Kommunikationsprotokolle TCP/IP) und bildet das technische Rahmenwerk für den Zugriff auf verknüpfte Dokumente, die innerhalb der vernetzten Computersysteme des Internets verteilt sind. Andere bekannte Dienste sind z.B. der E-Mail-Dienst und der FTP-Dienst (FTP steht für File Transfer Protokoll). Die Dokumente können mit einer grafischen Benutzeroberfläche, dem sogenannten Web- bzw. Internet-Browser, abgerufen werden. Die Geschichte des WWW begann 1989 am CERN (europäisches Zentrum für Atomphysik) in Genf. Hier wünschten sich die an den hochkomplexen Forschungsversuchen beteiligten Wissenschaftler eine einfache und schnelle Möglichkeit des Dokumentenaustauschs, da die Wissenschaftler geografisch weit verstreut zusammen an den Projekten arbeiteten. Der erste konkrete Vorschlag kam 1989 von dem CERN-Physiker „Tim Bernes-Lee“ und konnte dann 18 Monate später als Prototyp in Betrieb genommen werden. Noch war die Technik zu 100 Prozent textorientiert, was aber damals völlig ausreichte. 1993 konnte schließlich die erste Version der Weltöffentlichkeit mit der ersten grafischen Benutzeroberfläche „Mosaic“ präsentiert werden. 1994 wurde dann das WWW-Consortium (www.w3.org) zur Weiterentwicklung des Web gegründet.

Wie funktioniert das WWW?

Technisch gesehen basiert das WWW auf einer „Client/Server-Technologie“, d.h. der Server stellt einen Dienst zur Verfügung (hier den WWW-Dienst), den ein Client anfordert. Im Wesentlichen sind drei Bausteine entscheidend und dienen auch der webbasierten Gestaltung verteilter Anwendungssysteme:

- das HTTP-Protokoll (Hypertext Transfer Protocol) dient zum Übertragen von Informationen (Repräsentationen von Ressourcen in den Dokumenten). Da ein Web-Server mit diesem Protokoll kommuniziert, nennt man einen Web-Server auch HTTP-Server.
- ein Adressierungsschema (ist standardisiert) „URL“ (Uniform Resource Locator) als Spezialisierung von URI (Uniform Resource Identifier) für einen globalen Adress- und Namensraum.
- standardisierte Formate bzw. formale Auszeichnungssprachen (X)HTML (EXtensible Hypertext Markup Language) für die Repräsentation der Dokumente. HTML repräsentiert die Mensch-zu-Maschine-Kommunikation auf Basis von XML (EXtensible Markup Language), die die Maschine-zu-Maschine-Kommunikation repräsentiert. Einfach ausgedrückt: Die Webseiten werden mit der Beschreibungssprache (X)HTML beschrieben (später mehr dazu) (siehe Abb. 1).

Die Client-Seite

Aus der Sicht des Dienst-Nutzers besteht das Web aus einer mittlerweile gigantischen Sammlung von Dokumenten, die auch als Webseiten oder „Webpages“ bezeichnet werden. Jede Webseite kann „Links“ zu anderen Webseiten beinhalten, die wiederum andere verlinken können. Mit der Maus kann der Dienst-Nutzer den markierten Link anklicken und so auf die verlinkten Webseiten gelangen. Webseiten, die auf andere Webseiten verlinken, basieren auf dem sogenannten „Hypertext“. Als Hypertext bezeichnet man Texte, die mithilfe einer netzartigen Struktur von Objekten Informationen durch Hyperlinks zwischen Knoten (Wissenseinheiten) verknüpfen. (Quelle: Wikipedia) Die Webseiten können mit einem „Web-Browser“ angezeigt werden. Der Web-Browser stellt den Client dar und

fordert eine Webseite vom Web-Server an, indem der Benutzer die Adresse der Webseite in das Adressfeld des Web-Browsers einträgt. Der Web-Server bekommt die Anfrage mitgeteilt und transferiert die gewünschte Webseite dann zum Web-Browser. Man kann unschwer erkennen, dass der WWW-Dienst im Hintergrund auf einem Nachrichten-Protokoll basiert, dem HTTP-Protokoll. Das HTTP-Protokoll überträgt also eine angeforderte Webseite vom Web-Server zum Web-Browser. Dabei wird die ganze Webseite in eine HTTP-Nachricht verpackt und dann entweder verschlüsselt oder unverschlüsselt zum Web-Browser übertragen. Da immer häufiger auch andere Medien wie z.B. Videoclips, Soundtracks und Grafiken in die Webseiten integriert werden, muss auch die Datenübertragungstechnik mit der Weiterentwicklung der Web-Technologie Schritt halten. Die Hypertext-Webseiten werden immer mehr zu „Hypermedia“-Webseiten ausgebaut und benötigen superschnelle Datentransferraten (z.B. VDSL-Technologie). Neben den klassischen Webseiten werden im Web immer häufiger sogenannte „Web-Anwendungen“ eingesetzt. Web-Anwendungen sind im Gegensatz zu reinen Webseiten ganze und teils sehr mächtige Programme wie z.B. Google-Docs (Office-Anwendung), die komplett im Web-Browser des Benutzers ablaufen. Der Benutzer benötigt hierfür keine eigene Software auf dem heimischen Computer. Die Web-Anwendung läuft auf einem Web-Server des Anbieters und wird auch automatisch immer auf die neueste Version gebracht. Der Vorteil liegt auf der Hand: Der Benutzer oder Benutzergruppen sparen sich Kosten und Zeit, da ja alles der Dienstanbieter übernimmt.

Web-Browser im Überblick

Folgende Auflistung gibt einen Überblick über die derzeit gängigen kostenfreien

Web-Browser mit den dazugehörigen Internetadressen:

- Mozilla Firefox**
http://www.mozilla-europe.org/de/firefox/
- MS Internet-Explorer**
http://www.microsoft.com/germany/windows/internet-explorer/
- Opera**
http://www.opera.com/
- Safari**
http://www.apple.com/safari/
- Google Chrome**
http://www.google.de/chrome
- Seamonkey**
http://www.seamonkey-project.org/
- Konquerer (nur Linux)**
http://www.konqueror.org/

Die Server-Seite

Wie bereits erwähnt, nimmt der Web- bzw. HTTP-Server Anfragen von den Web-Browsern (Clients) entgegen, um dann die Webseite an den Client mittels HTTP-Protokoll herunterzuladen. Das bedeutet, dass der Web-Server die Webseiten bzw. Webpages auf seiner Festplatte speichert. Jede Webseite muss mit der Auszeichnungssprache HTML programmiert werden. Eine Webseite ist prinzipiell eine Datei, die z.B. so heißen könnte: „name_der_seite.html“.

Wenn ein Unternehmen eine Internetpräsenz besitzt, so existieren in der Regel mehrere Webseiten, die alle auf einem Web-Server eines Providers gespeichert werden müssen. Alle Webseiten zusammen nennt man „Website“. Technisch gesehen hat jede Website einen Server-Prozess, der den TCP-Port mit der Nummer 80 auf ankommende Verbindungen von Clients (Web-Browser) abtastet. TCP bedeutet „Transmission Control Protocol“ und ist eine Vereinbarung (Protokoll) darüber, wie Daten zwischen Rechner-Systemen ausgetauscht werden. TCP in Verbindung mit IP (Internet Protocol) ist die Grundlage des Internets. Das IP-Protokoll ist für die Vermittlung, also „wie kommen die Daten zum Zielrechner“ zuständig. Jetzt stellt sich sofort die Frage, über welchen Mechanismus wird eine Webseite in einer extrem kurzen Zeit gefunden und im Web-Browser angezeigt? Genau dafür ist das Routing mit dem wichtigsten Internet-Dienst „DNS“ (Domain Name Service) zuständig. Im Vergleich zum Telefonieren ermittelt das DNS mit dem im Web-Browser vom Benutzer eingegebene Hostnamen bzw. Domainnamen (z.B. www.meinefirma.de) die dazugehörige weltweit einmalige IP-Adresse. Die IP-Adresse ist die Anschlussnummer im Internet und könnte so aussehen: „178.0.34.98“. Das DNS ist ein weltweit auf Tausende von Servern verteilter hierarchischer

```

<!-- ##### -->
<!-- # Datei: grundgeruest_htmldatei.html # -->
<!-- ##### -->

<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>Das ist der Titel der Seite</title>
</head>

<body>
<!-- ##### -->
<!-- # Hier steht ein Beispiel-Inhalt # -->
<!-- ##### -->
<h1>Überschrift</h1>

<p>
  Hier steht der Text des ersten Abschnittes.
</p>

<p>
  Hier steht der Text des zweiten Abschnittes.
</p>
<!-- ##### -->
<!-- # Hier endet der Inhalt # -->
<!-- ##### -->
</body>
</html>
    
```

Abb. 2: Grundgerüst einer (X)HTML-Datei.

Verzeichnisdienst, der den Namensraum des Internets verwaltet. Das DNS ist also für die Umsetzung von Domainnamen in IP-Adressen zuständig. Im Folgenden sollen die Schritte vom Mausklick des Benutzers im Web-Browser bis zur Anzeige der Webseite im Web-Browser aufgezeigt werden. Der Benutzer trägt folgende Adresse bzw. URL in das Adressfeld des Web-Browsers ein: *http://www.meinefirma.de/produkt1.html*

Die URL besteht aus drei Teilen:

1. Name des Protokolls. „http“ steht für das HTTP-Protokoll
2. Name der Maschine, auf der sich die Webseite befindet, also „www.meinefirma.de“
3. Name der Datei, also „produkt1.html“

Nun die einzelnen Schritte:

1. Der Benutzer trägt im Adressfeld den Domainnamen ein: z.B. www.burgardsoft.de
2. Der Web-Browser ermittelt die URL und fragt DNS nach der entsprechenden IP-Adresse
3. Der DNS-Dienst gibt als Antwort z.B. „178.0.34.98“ an den Web-Browser zurück
4. Der Web-Browser baut eine TCP-Verbindung zum Port 80 des HTTP-Server-Prozesses des Web-Servers auf
5. Der Web-Browser sendet einen sogenannten „GET-Befehl“ /produkt1.html
6. Der Web-Server www.meinefirma.de sendet die Datei „produkt1.html“ zum Client
7. Die TCP-Verbindung wird nun wieder getrennt
8. Der Web-Browser zeigt nun das Dokument, also die Webseite mit dem gesamten Inhalt an.

Die Auszeichnungssprache (X)HTML

(X)HTML ist, wie bereits erwähnt, eine Auszeichnungssprache, auch Markup-Sprache genannt, zur Beschreibung, wie die Dokumente formatiert sind. Durch das Einbetten der standardisierten Markup- bzw. HTML-Befehle in die HTML-Datei kann jeder Web-Browser jede Webseite korrekt lesen. Die HTML-Sprache ist einer ständigen Weiterentwicklung unterworfen und ist unglaublich komplex. Jede korrekt angelegte Webseite bzw. HTML-Seite besteht aus einem Head (<head> und </head>) und einem Body (<body und </body>) zwischen den HTML-Tags „<html>“ und „</html>“. Der HTML-Tag „</html>“ bedeutet hier das Ende der HTML-Datei. Jeder Beginn-Tag muss mit einem Ende-Tag abgeschlossen werden. In Abbildung 2 soll das Grundgerüst einer (X)HTML-Datei namens „grundgeruest_htmldatei.html“ gezeigt werden.

Jede korrekt angelegte Webseite bzw. HTML-Seite besteht aus einem Head (<head> und </head>) und einem Body (<body und </body>) zwischen den HTML-Tags „<html>“ und „</html>“. Der HTML-Tag „</html>“ bedeutet hier das Ende der HTML-Datei. Jeder Beginn-Tag muss mit einem Ende-Tag abgeschlossen werden. In Abbildung 2 soll das Grundgerüst einer (X)HTML-Datei namens „grundgeruest_htmldatei.html“ gezeigt werden.

Eine umfangreiche Einführung in die HTML-Sprache bietet die Website <http://de.selfhtml.org/>

ZT Der Autor



Autor Thomas Burgard entwickelt Applikationssoftware und professionelle Internetauftritte für Unternehmen.

ZT Adresse

Thomas Burgard Softwareentwicklung & Webdesign
Dipl.-Ing. (FH) Thomas Burgard
Bavariastr. 18b
80336 München
Tel.: 0 89/54 07 07-10
Fax: 0 89/54 07 07-11
E-Mail: info@burgardsoft.de
www.burgardsoft.de

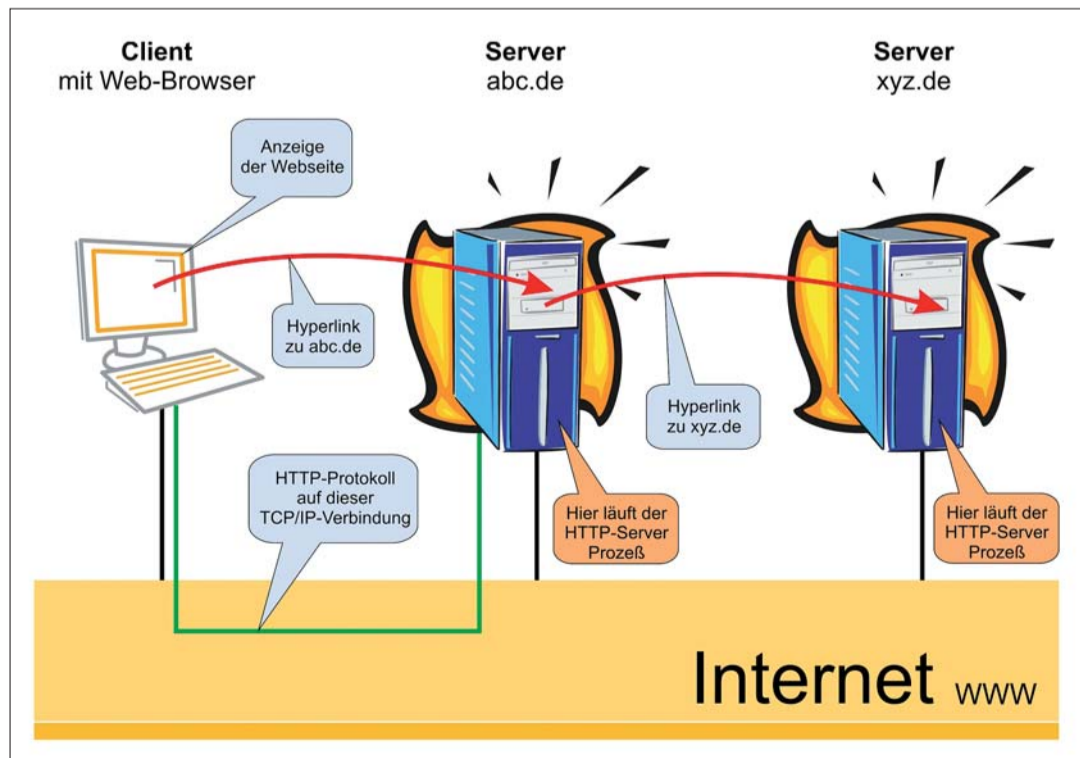


Abb. 1: web-architektur.jpg: Die Web-Architektur im Internet.