

1.1 HTML/Einführung

1.1.1 Was ist Hypertext? Wie kommt man auf so etwas?

Bis zum Ende der 80er Jahre waren digitale Informationen vorwiegend “am Stück” zu haben, d.h. mit Hilfe von Datenträgern, die individuell übergeben oder per Post verschickt wurden. Manche Vereine machten es sich z.B. zur Aufgabe, Open-Source-Software bekanntzumachen und zu verteilen.

Dann kamen Mailboxen auf und mit ihnen die Möglichkeit, Dateien zur Echtzeit einzustellen und abzurufen sowie E-Mails abzuschicken. Für beide Funktionen gab es eine einheitliche Oberfläche, die der Betreiber der Mailbox zur Verfügung stellte.

Zu Beginn der 90er Jahre begann dann das Internet, sich von einem reinen Forschungs- und Wissenschaftsnetz zu einem Rechnernetz für alle zu entwickeln. Die Menge an Informationen, die online verfügbar war, wuchs schnell an. Allerdings verteilte sie sich auf verschiedene Quellen, die über verschiedene Protokolle verfügbar waren:

- Dateiversand per FTP-Protokoll
- Eingang in Datenbank-Programme per TELNET-Protokoll
- E-Mail-Versand per SMTP-Protokoll, Empfang per POP3-Protokoll
- Diskussionen auf schwarzen Bretter per *usenet news* (NNTP-Protokoll)
- Online-Chat per IRC-Protokoll

Jede dieser Anwendungen benötigte ein eigenes Client-Programm mit eigener Oberfläche, häufig noch Konsolen-orientiert.

Wollte man sich Software herunterladen, musste man die Adresse des FTP-Servers wissen und sich dann per `cd`-Befehl weiterklicken. Für einen schnelleren Zugriff brauchte man die Adresse der zu holenden Datei selbst. Hier ein Beispiel aus dem Jahr 2012:

```
Terminal
schueler@debian964:~$ ftp ftp.uni-stuttgart.de
Connected to ftp.uni-stuttgart.de.
220 ProFTPD 1.3.2c Server (ftp.uni-stuttgart.de) [129.69.8.112]
Name (ftp.uni-stuttgart.de:schueler): anonymous
331 Anonymous login ok, send complete email address as your password
Password:
230 Anonymous access granted, restrictions apply
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> cd pub
250 CWD command successful
ftp> ls
200 PORT command successful
150 Opening ASCII mode data connection for file list
drwxr-xr-x  3 553          root    4096 Feb 16 2010 download
drwxr-xr-x 21 ftpmirror  nogroup  4096 Apr 12 2012 mirrors
drwxrwsr-x  5 sys        1005    4096 May 11 2011 systems
226 Transfer complete
ftp> cd download
250 CWD command successful
ftp> ls
200 PORT command successful
150 Opening ASCII mode data connection for file list
-rw-r--r-- 1 553  143  4935804 Mar 28 2007 BeispielMediaVorlesung.mp4
lrwxrwxrwx 1 553  143           28 Feb 15 2008 Release_3_9_ecs_branch.tgz
```

```
drwxrwxr-x 2 1031 root    4096 May  6 2009 num_meth
226 Transfer complete
ftp> get BeispielMediaVorlesung.mp4
local: BeispielMediaVorlesung.mp4 remote:
BeispielMediaVorlesung.mp4
200 PORT command successful
150 Opening BINARY mode data connection for BeispielMediaVorlesung.mp4
(4935804 bytes)
226 Transfer complete
4935804 bytes received in 17.85 secs (270.1 kB/s)
ftp> bye
```

Eine so umständlich und schwierig erscheinende Oberfläche setzte der Ausbreitung bei Nicht-Fachleuten enge Grenzen. Zwei Lösungen boten sich an:

- a) *gopher* als hierarchisches System — hier muss man sich quasi von oben her zu einem bestimmten Thema auf einem (Gopher-) Server durchklicken. Texte zum GNU-C-Compiler fand man beispielsweise über den Weg:

```
Alles->Wissensch.->Informatik->Programmierung->C->Compiler->GNU-C
```

Da ein Teil der benötigten Software kostenpflichtig war, konnte sich dieser Ansatz trotz des großen anfänglichen Vorsprungs nicht behaupten. Heute spielt er keine Rolle mehr.

- b) *www* als unstrukturiertes System — in der Wissenschaft, aber auch in Zeitschriften wie *c't* und *iX*, war es seit jeher üblich, dass Artikel Zitate, Quellenverweise, Literaturlisten und Verweise auf andere Internet-Quellen enthalten können. Nun musste man nur noch das Folgen dieser Verweise (Links) automatisieren.

Für diese Lösung sind als Voraussetzung nötig:

- Die Datei enthält Verweise in maschinenlesbarer Form: Hier wurde das HTML-Format erfunden¹
- Der Dateileser ist in der Lage, diesen Verweisen auf Knopfdruck zu folgen (d.h. auf die verwiesene Datei zuzugreifen und sie ebenfalls anzuzeigen): Hierzu eignet sich ein Browser-Programm wie Firefox
- Es gibt einen Server, der (wie ein FTP-Server auch) auf Anfrage eine Datei herausgibt und ein Protokoll, das (wie das FTP-Protokoll auch) diese Dateiübertragung ermöglicht: Als Serverprogramm dient heute oft Apache; das Protokoll heißt HTTP².

Alle diese Voraussetzungen waren von Anfang an weitgehend frei verfügbar, so dass sich diese Lösung rasant ausbreitete.

Anders als beim gopher-Dienst gibt es beim WWW ursprünglich keine Möglichkeit, eine Information gezielt zu finden (es sei denn, man weiß schon, wo sie ist). Diese Lücke wird durch die Suchmaschinen schon seit Ende der 90er Jahre geschlossen. Später gab es Meta-Suchmaschinen, die ihrerseits andere Suchmaschinen befragten und deren Dienste ausnutzten.

¹HTML ist in seinen ersten Versionen eine Ausprägung des SGML-Formats, das in den 80er Jahren für die Speicherung von Buchtexten erfunden wurde

²Es ist ursprünglich sehr einfach aufgebaut und damit simpel zu programmieren.