

Inhaltsverzeichnis

1 Was ist das WWW?	2
1.1 Unterscheidung WWW – Internet	2
2 Frühgeschichte des WWW – Berners-Lee am CERN	3
2.1 Entwicklungsprozess nacherzählen	3
2.2 Was entwickelte Berners-Lee am CERN?	3
3 Durchsetzung des WWW (1992-1996)	4
3.1 neue Browser (Auswahl)	4
3.2 neue Strukturen und Adaption – das WWW als „success disaster“	4
4 Exkurs: alternative Informationssysteme	5
4.1 Hypertext ohne Internet	5
4.1.1 Memex – Vannevar Bush (1945)	5
4.1.2 oN Line System – Doug Engelbart (1968)	5
4.1.3 Xanadu – Ted Nelson (1965)	5
4.2 Internet ohne Hypertext – frühe 90er	6
4.2.1 WAIS – Brewster Kahle	6
4.2.2 Gopher – Mark McCahill et al.	6
4.3 Warum funktioniert das WWW so gut im Internet?	6
4.4 Warum konnte sich das WWW so gut im Internet entwickeln?	6
5 Wer reguliert das WWW?	6
5.1 W3C	6
6 Topologien des WWW	7
6.1.1 Wachstum und preferential attachment	8
7 Diskussion: Welche Auswirkungen haben diese Topologien?	9
7.1.1 Lessigs Szenario: Zertifikate und passive Nutzer	9
7.2 Diskussionsfragen	9

WWW, das success disaster

Welche Techniken wurden verworfen, bevor es zum WWW kam? Warum scheint das WWW so gut im Internet zu funktionieren?

1 Was ist das WWW?

Einführend:

- zuerst: Woher kommt das WWW? Entstehungsgeschichte (Märchenonkel) und einige Alternativen. Aber nicht: Vor Browserkrieg (ca. 1994) ist Schluss; keine aktuellen Standards.
- Anschließend: kurz auf WWW als Ganzes eingehen. Warum hat es sich durchgesetzt? Warum funktioniert es so gut im Internet? Wie funktioniert seine Regulierung?
- Danach Exkurs als Kontrast: Alternative Informationssysteme.
- Dann: Gesamtbild Welche Topologie(en) entwickelt (entwickeln) sich im Web?
- Diskussion: Was ergibt sich daraus? Was für ein Web wollen wir?

1.1 Unterscheidung WWW – Internet

- Internet = physische Infrastruktur
- WWW = Dienst im Internet (wie E-Mail, FTP),
 - Hypertext – System. Gerichtete Links. Nutzer kann Dokumente betrachten, aber nicht verändern.
 - Etwas verwirrend: Immer mehr Dienste, die ursprünglich vom WWW getrennt waren und mit getrennten Programmen genutzt werden, werden immer öfter über die technischen Möglichkeiten des WWW angeboten und können mittels eines Browsers genutzt werden. (z.B E-Mail als Webmail, Usenet über GoogleGroups...)
- Technisch gesehen, ist der Kern des WWW eine Sammlung aus:
 - **Programmen (Server, Client, Browser)**
 - ♦ Server: lagert und verschickt Daten
 - ♦ Client: fordert Daten beim Server an
 - ♦ Browser: Zeigt Dateien an
 - einem **Protokoll (HTTP)** – zum Abruf und Versand von Dateien
 - einer Beschreibungssprache für Hypertext-Seiten (**HTML**)
 - ♦ beschreibt Aufbau der Seite
 - ♦ erlaubt Integration anderer Dateitypen (Graphik, Animation, Java) über Links
 - und einer Konvention (**URL**) – zur Adressierung von Informationen
 - ♦ weist jeder Ressource einen eindeutigen Standort zu
 - ♦ Aufbau: <Protokoll>://<Dienst>.<Server><Verzeichnis><Datei>

Weiterentwicklungen:

- Kleine Programmschnipsel (Applets) in Websites integriert (Java)
- WWW als Interface / Oberfläche für verteilte Programme (PHP)
- Nutzer können Dokumente verändern: Wikis, Groupware

2 Frühgeschichte des WWW – Berners-Lee am CERN

- Warum am CERN?
 - Problem am CERN: Grundverschiedene Computersysteme, Software, Nutzungsgewohnheiten.
 - Bei jedem neuen Projekt: Streit darüber, welches Computersystem / -programm benutzt werden sollte
 - aber auch: wissenschaftliche Umgebung und Traditionen, d.h. Kultur der Offenheit von Forschungsergebnissen und ähnlichen Daten

2.1 Entwicklungsprozess nacherzählen

- „Ich kam zufällig zur rechten Zeit und mit den passenden Interessen und Neigungen, nachdem der Hypertext und das Internet ihre Volljährigkeit erreicht hatten. Die einzige Aufgabe, die mir blieb, war es, die beiden miteinander zu verheiraten.“ (Berners-Lee 1999, 19)
- Berners-Lee wollte Informationen so organisieren, wie unser Gehirn es tut: unstrukturiert, verknüpft. (statt hierarchisch)

2.1.1

Vorstellung TBL:

- **Idee der Universalität:** TBL und Kollegen wollten einen potentiell allumfassenden Informationsraum schaffen. Daraus ergaben sich die meisten technischen Entscheidungen.
 - Adressierung (URL)
 - Gleichbehandlung aller Dokumente: Der Client versucht nie herauszufinden, was ein Dokument bedeutet – er fordert es einfach an.
 - ♦ damit lassen sich auch Nicht-HTML-Dokumente einbinden – z.B. die vorhandenen riesigen FTP-Archive
 - Möglichkeit der Verknüpfung von Dokumenten, unabhängig von deren Ort.
 - Skalierbarkeit (wenn's im Kleinen funktioniert, klappt es auch im Großen)
 - Dezentralität (keiner kann so ein Riesenprojekt managen)
- Bis dahin waren Hypertextsysteme Datenbanken, die auf einer Festplatte oder Diskette lagerten und nur interne Verknüpfungen kannten.

2.2 Was entwickelte Berners-Lee am CERN?

- Enquire (1980): Verknüpfung von Karteikarten über Personen, Computer,
- Enquire lief nie in einem Netzwerk.
- RPC (1984): Programm ermöglicht es, Prozeduren auf einem entfernten Rechner aufzurufen

- Client-Browser-Editor: *WorldWideWeb*, fertig im Okt 1990
- dann HTTP, URL
- Dez. 1990: HTML als Sprache, die beschreibt, wie Seiten mit Hypertextverknüpfungen formatiert werden.

- Dez. 1990: Server-Software: speichert Webseiten auf einem Computer und erlaubt den Zugriff darauf.
 - richtet **ersten Server** auf seinem Computer ein: *info.cern.ch* Dort lagen – erste Webseite – die Spezifikationen von HTTP, URI und HTML sowie alle projektbezogenen Informationen.
- während 1991: Software auf anderen Plattformen zum Laufen bringen.
- Wichtige Schritte:
 - Einbindung FTP: Damit standen die Archive der Internet-Newsgroups als Content zur Verfügung
 - 1991: Einsatz des WWW als Telefonverzeichnis am CERN
 - Aug. 1991: TBL stellt Web-Projekt in der Newsgroup [alt.hypertext](#) vor. Stellt Client (für NeXT), Browser und Server-Software zur freien Verfügung, weil er mehr Programmierer braucht, um das Projekt zu verfolgen.
- daraufhin werden neue Server eingerichtet, der Anfang einer Web-Gemeinde entsteht.

3 Durchsetzung des WWW (1992-1996)

- 1992 gab es innovative Websites, z.B. ein virtuelles Museum und einen virtuellen Atlas
- Wissenschaftler und Regierungsgruppen (Publikationspflicht) stellen ihre Informationen ins WWW, statt sie auf Anfrage zu versenden --> mehr Leute besorgen sich Browser
- WWW-Software in Public Domain (30. April 1993)
 - Gopher als warnendes Beispiel
 - förderte Akzeptanz des WWW durch Firmen (IBM u.a.)
- Zugriffe/Tag auf info.cern.ch: Faktor 10 pro Jahr (Sommer 1993=10.000). Immer neue Server.
- Mitte 1994: Microsoft macht das Internet zum Teil der Firmenstrategie und bietet in Windows95 Software für den Internetzugang an.
 - Integration des WWW in die am stärksten verbreitete Plattform
- 1996: AOL und Microsoft einigen sich, AOL mit einer IE-Version auszuliefern. Damit kommt noch einmal eine große Nutzergruppe hinzu, die schon an den Umgang mit dem Web gewöhnt ist.

3.1 neue Browser (Auswahl)

nun begannen auch externe Programmierer, Browser zu entwickeln.

- ViolaWWW (für Unix Anf. 1992) – Grafiken, Animationen, Applets. Aber kompliziert zu installieren.
- Cello (Anf. 1993) für Windows / IBM-Kompatible (am stärksten verbreitete Plattform)
- Lynx (Anf. 1993) Nur Text, aber scrollbar (heute: oft von Blinden genutzt)
- Mosaic, später Netscape
 - Auf Benutzbarkeit und leichte Installation optimiert, für verschiedene Plattformen
 - quasi als Produkt vermarktet
 - Firma gegründet, in Netscape umbenannt. (April 1994)

3.2 neue Strukturen und Adaption – das WWW als „success disaster“

- Web wächst nicht zum Medium der Zusammenarbeit, sondern als Publikationsmedium (wg. fehlenden Editoren)

- viele Fragen traten erst mit der Ausbreitung des WWW auf (z.B. geistiges Eigentum im digitalen Bereich); es gibt immer neue Probleme; statt konsistenter Lösungen experimentiert jeder drauflos

4 Exkurs: alternative Informationssysteme

4.1 Hypertext ohne Internet

4.1.1 Memex – Vannevar Bush (1945)

- Idee: Den menschlichen Intellekt von Routineaufgaben (wie der Suche nach Information) entlasten, damit er sich auf die Schaffung von Wissen konstruieren kann.
- erste Idee eines Hypertext-Systems, in einen Schreibtisch integriert
- Dokumente sollten auf Mikrofilm gespeichert sein.
- Nutzer sollten Dokumente über „Trails“ miteinander verknüpfen, also Pfade zwischen ihnen anlegen.
 - Bush: „*The inheritance from the master becomes, not only his additions to the world's record, but for his disciples the entire scaffolding by which they were erected.*“ Bush sah also schon die Möglichkeit, die Struktur von Wissen durch Analyse des Hypertextes abzubilden.

4.1.2 oN Line System – Doug Engelbart (1968)

- Pionierarbeit: grundlegenden Elemente der Mensch-Computer-Interaktion, Gruppenarbeit in Netzwerken, die Implementierung von Hypertext und die Verwendung des Computers als Kommunikationsmedium
- Rechner als Medium zur Unterstützung menschlicher Denk- und Arbeitsprozesse – statt als Rechenknecht. (1962)
- NLS technisch:
 - Mainframe versorgt Terminals mit Rechenleistung
 - Terminal: Tastatur, Modus-Tastatur, Maus (!!!)
 - Grafische Benutzeroberfläche mit Fenstern. Daten durch grafische Symbole repräsentiert.
 - verschiedene „Dienste“: Schwarzes Brett, Messaging (ähnlich E-Mail), Textverarbeitung, Videokonferenz
- unglaublich fortgeschritten für 1968 – Maus und grafische Benutzeroberfläche schafften es erst 1984 in den Massenmarkt (Apple Macintosh)
- Nachteil: Hochkompliziert aufgebaut, schwer zu bedienen; teures High-Tech. Benutzer mussten eigene Befehlssprache lernen.

4.1.3 Xanadu – Ted Nelson (1965)

Soziologe und Philosoph („professioneller Visionär“)

- Nelson gilt als Erfinder des Begriffs und Konzepts „Hypertext“, 1965.
- schlug 1965 ein System namens Xanadu vor. Nutzung von Computern als Infrastruktur für eine neue Art des literarischen und künstlerischen Denkens und Produzierens
 - Eher geisteswissenschaftliche Herangehensweise; definierte sein System nicht von der Technik her (wie TBL), sondern von der Erweiterung der Möglichkeiten des menschlichen Intellektes. Außerdem war er dem Stand der Technik seinerzeit weit voraus.

→ „*Computer people don't understand computers. Oh, they understand the technicalities all right, but they don't understand the possibilities.*“ (Nelson 1999)

- Im WWW sieht Nelson nur eine partielle Umsetzung seiner Ideen („Papier-Simulator“). Möglichkeiten der Computer und Netzwerke nicht richtig ausgenutzt.

4.2 Internet ohne Hypertext – frühe 90er

Alternativen zum WWW. TBL integrierte Gopher und WAIS ins WWW, damit sie es nicht erdrückten.

4.2.1 WAIS – Brewster Kahle

- Netz aus Suchmaschinen – System zur Volltextsuche in verteilten Datenbeständen
- Suchsystem für Datenbanken

4.2.2 Gopher – Mark McCahill et al.

- unformatierter Text
- automatisch generierte Menüs über Inhalt des Servers
- Suchsystem: Veronica
- Trotz relativ weiter Verbreitung scheiterte Gopher. Dies lag am Versuch der Universität von Minnesota, die Nutzung der entsprechenden Programme kostenpflichtig zu machen.
 - Warnendes Beispiel für TBL; daher Antrag: WWW-Software in Public Domain.

4.3 Warum funktioniert das WWW so gut im Internet?

- Offene Standards . Public Domain - Software
- minimalistisches Designprinzip
 - Hardware-Unabhängigkeit
 - viele Einsatzmöglichkeiten (Einbindung verschiedener Dateitypen etc.)
- von TBL und späteren Entwicklern entlang konkreter Bedürfnisse entwickelt, statt als Vision (Kompromissbereitschaft)
 - Bedingung für Aufnahme des Projektes durch andere Entwickler
- kam zur rechten Zeit; andere Systeme waren den technischen Möglichkeiten ihrer Zeit zu weit voraus. Das WWW entstand dagegen auf Basis vorhandener Technik

4.4 Warum konnte sich das WWW so gut im Internet entwickeln?

- nach Barabási: info.cern.ch als Hub; Modularität begünstigt Multitasking
- eignete sich besonders gut für diese Art der Entwicklung (plattform-unabhängig, nichtmateriell)
 - nur so konnten die nötigen Ressourcen zusammenkommen

5 Wer reguliert das WWW?

5.1 W3C

- Mitte 1994 am MIT unter Leitung von TBL gegründet.

- zunächst gemeinsam von MIT und CERN getragen, CERN stieg später aus.
- seit ca. 1996 dritte Wirtsinstitution: Keio-Universität in Japan
- Ziel: „leading the Web to its full potential“
 - technische Spezifikationen entwickeln
 - Vertretung von Entwicklern, Wissenschaftlern und Benutzern
- Mitglieder: Bildungs- und Regierungsorganisationen, Unternehmen.
- Arbeitsweise:
 - Mitglieder machen Vorschläge
 - es bildet sich eine Arbeitsgruppe
 - Vorschläge bearbeitet, zurück an Mitglieder; Prozess kann sich wiederholen
 - W3C gibt **Empfehlungen (keine Standards!)**
- Arbeitsweise des W3C reflektiert einige Möglichkeiten und Grenzen der Kontrolle im Web.
 - Möglichkeiten: Wenn sich genug einflussreiche Gruppen für eine Empfehlung aussprechen, hat sie Aussicht auf großflächige Umsetzung.
 - Begrenzungen: Andererseits ist eine vollständige Umsetzung einer solchen Empfehlung nicht zu erwarten.
 - Das W3C kann z.B. die breitestmögliche Lesbarkeit einer Website als Anreiz bieten, sie in gutem HTML zu schreiben; aber es kann es nicht erzwingen.

6 Topologien des WWW

- Elemente der WWW-Topologie: Selbstorganisation, Wachstum, Dezentralität, gerichtete Links.
 - Skalenfreiheit (Potenzgesetze) als Anzeichen selbstorganisierender Netzwerke
- WWW als komplexes, sich entwickelndes System
 - eher strukturiert wie ein Ökosystem oder eine Zelle als wie eine Maschine
- **Frage:** Welche **Strukturen** bilden sich heraus, wenn die wohlbekanntesten Einzelteile des Web (oder des Internet) zusammenspielen?

Hierarchisch-modularer Aufbau:

- Modular: viele kleine, intern dicht verlinkte Module ordnen sich zu größeren, weniger dicht verlinkte Clustern (die sich wiederum zusammenschließen etc...)
- Konsequenz: unterschiedliche Module können getrennt voneinander arbeiten und sich dennoch koordinieren. (Multitasking)
 - Nachteil: Beim Multitasking kann ein Engpaß am koordinierenden Hub entstehen.
- **Kontinente** (S. 166 ff.) als notwendige Folge der gerichteten Links
 - In-Kontinent und Inseln können von Suchmaschinen nicht erfasst werden
 - Deshalb sind nur etwa die Hälfte der Web-Dokumente durch Surfen erreichbar.
 - (nämlich bei optimalem Startpunkt Central Core und Out-Kontinent; als Ausgleich bleibt der allergrößte Teil des Web für jemanden unerreichbar, der auf einer Insel surft – also durchschn. nur ein Viertel erreichbar)
- **Communities** bilden sozusagen die Städte und Dörfer des Webs: Kleine, intern stark verlinkte Gemeinden.

- keine klare Def. für Community. Vorschlag: Eine C. wird von Webseiten gebildet, die mehr Links zueinander als zu Seiten außerhalb haben.
- Das small world-Phänomen entsteht v.a. durch Sites, die viele Out-Links versammeln (Index-Seiten, Kataloge / Directories). Sie bieten eine Abkürzung zwischen entfernten Sites.
- Wachstum und Reifung des WWW:
 - das WWW bleibt trotz seines Wachstums skalenfrei. (74), es wird also keine charakteristische Seitengröße entwickeln
 - theoretisch mögliche Endstadien des WWW:
 - bleibt skalenfrei, solange das Wachstum groß genug ist
 - vollständig verbunden (alle Knoten mit allen anderen verknüpft)
 - Reifung (ähnlich der Bose-Einstein-Kondensation: Einige -letztlich ein einziger- Hubs vereinen alle Links auf sich.) Tritt ein: bei wenig Wachstum und vielen neuen Links (wg. preferential attachment)

6.1.1 Wachstum und preferential attachment

einige wenige Sites ziehen den Großteil der Nutzer auf sich

<i>Verteilung des Nutzervolumens zwischen allen Sites, Erotik-Sites und .edu-sites (individuelle AOL-Nutzer am 1. Dezember 1997)</i>			
% des Nutzervolumens	% der Sites		
<i>(=top x% der Sites)</i>	<i>alle Sites</i>	<i>Erotik-Sites</i>	<i>.edu-Sites</i>
0,1	32,36	1,4	2,81
1	55,63	15,83	23,76
5	74,81	41,75	59,5
10	82,26	59,29	74,48
50	94,92	90,76	96,88

Quelle: Adamic/Huberman 2001

- Wie können neue Websites alte überholen? Einige Websites wachsen sehr viel schneller als andere. Wie kommen diese Unterschiede zustande?
- Modell: Netzwerke WWW, Nutzer übereinander legen
 - Verbindungen zwischen WWW-Seiten (Links)
 - bereits stark verlinkte Sites werden häufiger besucht und erhalten mehr Links
 - Verbindungen zwischen Menschen (Mundpropaganda)
 - Nutzer hören eher von populären Sites als von unbeliebten.
 - zweimal preferential attachment
- Wirtschaftliche Erklärung: **increasing returns**.
 - Anders als in den meisten Wirtschaftszweigen, gilt für Internet-Seiten (aber auch Software-Industrie, Fernsehsender) das Gesetz der increasing returns: Je größer, desto rentabler.
 - können also auch mehr Geld für Werbung ausgeben, sich das Netz der Mundpropaganda zunutze machen.

- Ergebnis: Doppelter Vorsprung (stärker verlinkt und bekannter)
- so führen Bevorzugung der meistverlinkten Seiten und Mundpropaganda zum außerordentlich schnellen Wachstum einiger Websites - und eben nicht allein deren Qualität

7 Diskussion: Welche Auswirkungen haben diese Topologien?

Frage der Regulierbarkeit: kurz Lessigs Szenario vorstellen, dann diskutieren

7.1.1 Lessigs Szenario: Zertifikate und passive Nutzer

- nach Lessig:
 - Cyberspace nicht „von Natur aus“ unregulierbar – hat keine Natur
 - Regulierung möglich über Zusammenspiel von Handel und Regierung. Regierung reguliert Handel, Handel setzt Anreize für Nutzer; so können neue Standards durchgesetzt werden.
 - oder Regulierung über Verpflichtung von Software- und Hardware-Firmen
 - Gegenmittel: Open Source (Code demokratisch und modular erstellen)
- E-Commerce als treibender Faktor für Regulierbarkeit:
 - Handel ist an sicherer Umgebung interessiert. Dazu gehört die Identifizierbarkeit der Käufer. Also wird der Handel den Einsatz digitaler IDs fördern.
 - E-Commerce kann dann Anreize für die Käufer setzen, sich eine digitale ID zuzulegen; so ließe sich das Prinzip großflächig durchsetzen. („certificate-rich world“)
 - dann: ID als Ansatzpunkt für Regulierung durch Politik. Nicht perfekte, aber effektive Kontrolle.

Lessigs Tenor:

- Wir können entscheiden, welcher Code den Cyberspace regulieren soll. Wenn wir uns passiv verhalten, wird diese Entscheidung entsprechend den Bedürfnissen des Handels getroffen werden. Lessig sieht die meisten Menschen als passiv („cattle“)
- Lessig wünscht sich eine Art „Umweltbewegung“ für den Cyberspace, die eine offene und demokratische Informationspolitik einfordert. (Y2K-Problem als „code-based environmental disaster“)

7.2 Diskussionsfragen

- Stimmt das Szenario? Muss es so eintreten? Wenn nein, warum nicht?
- Welchen Cyberspace wollen wir?
- Was können wir dafür tun?

Kritik an Lessig:

- operiert mit unpräzisen Begriffen (Code, Architektur, Cyberspace) Kritik Barabási an Lessig: Code sind nur die Bausteine. Architektur ist, was wir aus ihnen errichten. Code determiniert nicht die Architektur.
- seine Vision des Cyberspace ist an der realen Welt modelliert. Deshalb lässt sie die anderen Möglichkeiten im digitalen Raum außer acht (z.B. multiple IDs, Anonymisierung...)
- Lessig vereinfacht: er sieht den Cyberspace als einheitlichen Bereich, in dem jede Änderung alle gleichermaßen betrifft

- Meine These – zur Diskussion: Auf der Inhaltsebene ist das Web tatsächlich kaum regulierbar. Es müssen schon andere Elemente hinzukommen.
 - so versagen national orientierte Rechtssysteme angesichts eines internationalen Informationsverbundes
- von wem kann Regulierung ausgehen?
 - Nutzer (Bsp.: Eltern setzen Filter für ihre Kinder)
 - können den eigenen privaten Bereich regulieren
 - können durch ihre Entscheidungen (Software, Hardware, Online-Verhalten) bestimmte Methoden oder Kulturen beeinflussen
 - Regierungen:
 - können bestimmte Inhalte auf inländischen Servern verbieten
 - können Offenlegung von ISP-Daten, E-Mails etc. verlangen
 - können auf einzelne kommerzielle Akteure zugreifen (Microsoft, Yahoo!, AOL, Telekom...)
 - können anonymen Internet-Zugang erschweren
 - Hardware-Hersteller:
 - können Computer durch Hardware-Codes identifizierbar machen (Trusted Computing)
 - können entscheiden, bestimmte Geräte einzubauen oder nicht
- These: Dies sind nur Einflüsse, keine harten Regeln. Aber sie können die Richtung der Entwicklung mitsteuern. Die immer wieder neu gefällte Entscheidung über die Richtung des WWW ist ein politischer Prozess.