

Hellmut Riediger

Recherchieren: Grundsätze und Grundbegriffe



Einführung in die Internetrecherche

Geschichte, Struktur und Allgemeines



Geschichte: Internet, WWW, Google

- 1969 Gründung von Arpanet, dem Vorläufer des Internets: die ersten vier Großrechner in der UCLA, im SRI, der University of California in Santa Barbara (und der University of Utah) werden miteinander verbunden
- 1989 Erfindung des WWW durch Tim Berners-Lee am CERN in Genf
- 1991 von CERN aus geht das von Tim Berners-Lee entwickelte World-Wide-Web (WWW) ins Netz
- 1998 Gründung von Google
- 2015: http://media.aperto.de/google_epic2015_de.html

Merkmale des WWW

Vom Text zum Hypertext

Das WWW basiert auf drei Kernstandards:

- HTTP als Protokoll, mit dem der Browser Informationen vom Webserver anfordern kann.
- HTML als Dokumentbeschreibungssprache, die festlegt, wie die Information gegliedert ist und wie die Dokumente verknüpft sind (Hyperlinks).
- URLs als eindeutige Adresse bzw. Bezeichnung einer Ressource (z. B. einer Webseite), die in Hyperlinks verwendet wird.

Wie groß ist das WWW? Anzahl Webseiten

- Dezember 1990 1
<http://info.cern.ch/hypertext/WWW/TheProject.html>

(<http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html>)

- Dezember 1994 10.022
- Januar 1996 100,000
- Dezember 2000 25,675,581
- Dezember 2003 45,980,112
- November 2006 101,435,253

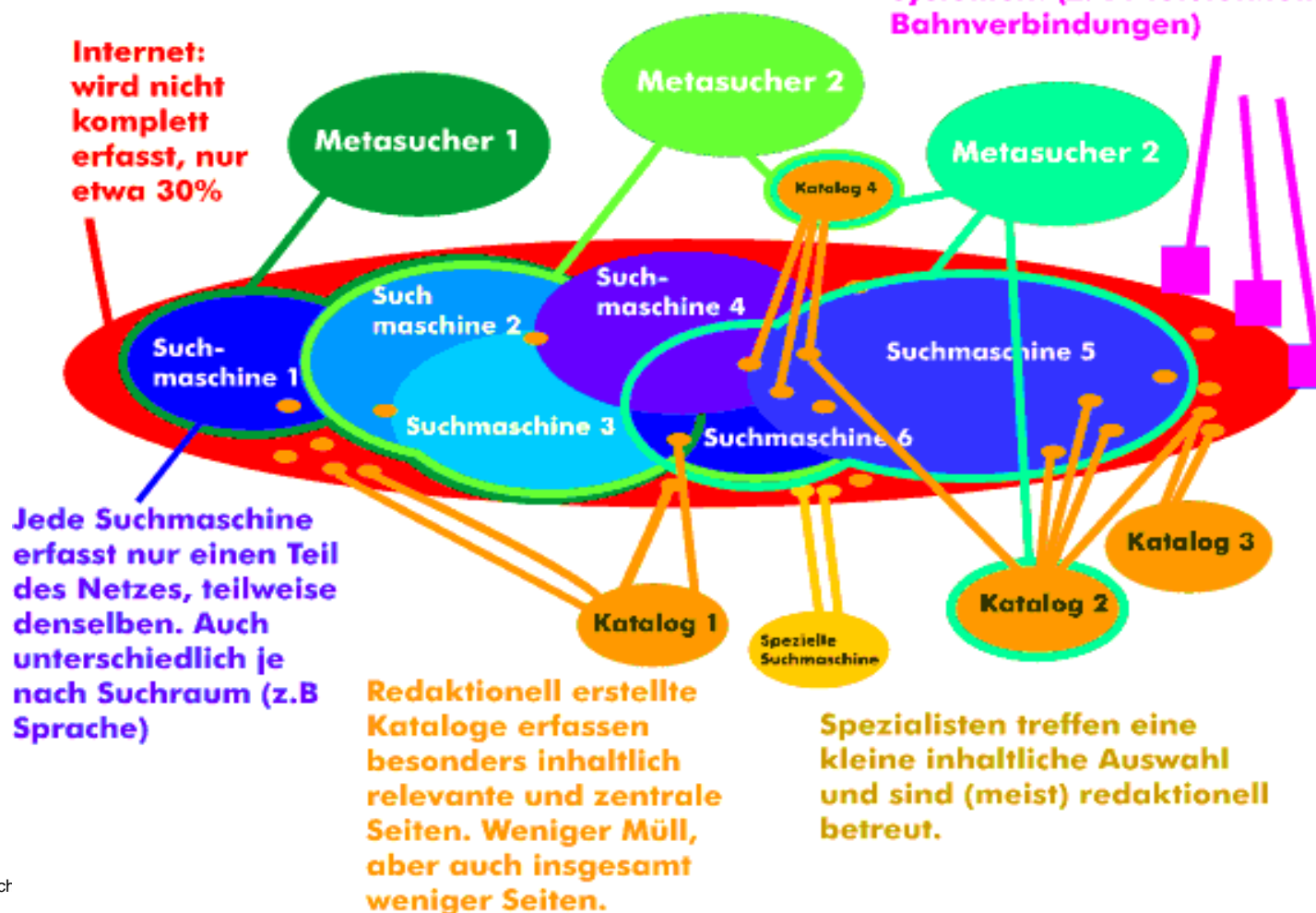
(Quelle: <http://www.zakon.org/robert/internet/timeline>)

„Suchraum“

Quelle: <http://www.suchfibel.de/2kunst/suchraum.htm>

Jeder Metasucher durchsucht mehrere Suchmaschinen und auch Kataloge. Die Datenbasis wird damit größer. Aber auch der Datenmüll.

Datenbanken mit speziellen Informationen. Lediglich der Zugriff funktioniert über das Internet, die Daten selber stammen aus unabhängigen Systemen. (z. B. Telefonnummern, Bahnverbindungen)



Das unsichtbare Web

„Das unsichtbare Web“
(„Deep Web“ bzw. „Invisible Web“): viele
Dokumente können von Suchmaschinen nicht
gefunden werden, z.B.:

- dynamische Seiten (Datenbanken)
- zugriffsgeschützte Dokumente
- nicht verlinkte Seiten
- noch nicht indexierte Seiten
- durch Meta-Tag geschützte Seiten(robots.txt)

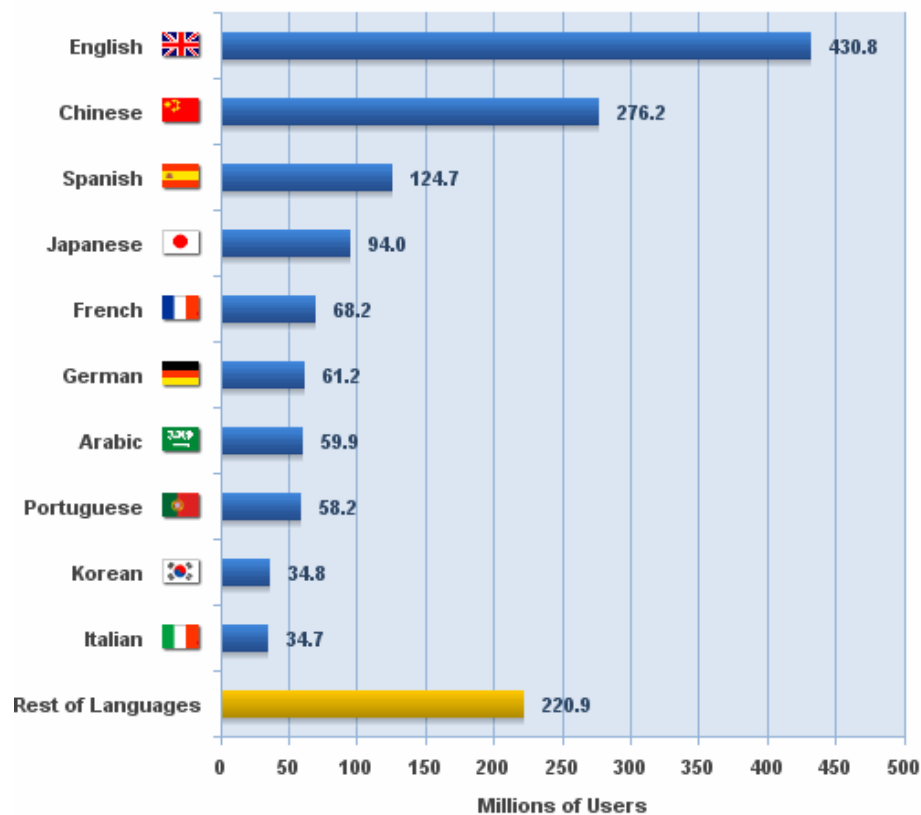
Infos zum „Deep Web“:

http://www.suchfibel.de/2kunst/das_unsichtbare_netz.htm

Sprachen im Web

1,463,632,361 Milliarden Internet-Nutzer (30. August 2008)

Top 10 Languages in the Internet



Source: Internet World Stats - www.internetworldstats.com/stats7.htm
Estimated Internet users 1,463,632,361 for 2Q 2008
Copyright © 2008, Miniwatts Marketing Group

Web 2.0

Web 2.0 ist ein Begriff für eine Reihe interaktiver und kollaborativer Phänomene des Internets, speziell des WWW.

Vgl.: <http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

Zum Beispiel:

- Wikis
- Weblogs
- Bild- und Videoportale (z.B. YouTube)
- „Socialbookmarking“ (z.B. <http://del.icio.us/>)
- Tauschbörsen (Peer-to-Peer- oder File-Sharing-Dienste genannt)
- Kollaborative Enzyklopädien und Wörterbücher (z.B. Wikipedia)
- Diskussionsforen (z.B. <http://www.proz.com/>)

Google-Daten und -Zahlen

- seit 1998 im Netz
- Börsenstart 2004
- gegründet von Mathematikern Sergey Brin und Larry Page
- 200.000 Computer für den Index (das größte Netz an Computern weltweit)
- Ca. 16 Milliarden Web-Seiten (??) indiziert
- 10,604 Milliarden \$ Umsatz im Jahr 2006

Wie groß ist der Google-Index (Anzahl Webseiten) ?

■ Google Oktober 2000	1.060.000.000
■ Google September 2002	2.469.940.685
■ Google Mai 2004	4.285.199.774
■ Google März 2005	8.058.044.651
■ Google (25.) September 2005	8.168.684.336
■ Google September 2008	20.000.000.000??

Google-Dienste



Search Services

- ▶ [Advanced Search](#) ▶ [Google Blog Search](#) ▶ [Google Book Search](#) ▶ [Catalogs](#) ▶ [Google Directory](#) ▶ [Google Earth](#) ▶ [Froogle](#) ▶ [Google Glossary](#) ▶ [Google Groups](#)
- ▶ [Google Image Search](#) ▶ [Google Library](#) ▶ [Google Local](#) ▶ [Google Maps](#) ▶ [Google Movies](#) ▶ [Google News](#) ▶ [Google Ride Finder](#) ▶ [Google Scholar](#) ▶ [Google Sets](#)
- ▶ [Travel Conditions](#) ▶ [Google Weather](#) ▶ [Google Video](#) ▶ [Google Web Search](#)

Specialized Searches

- ▶ [Apple-Specific Search Engine](#) ▶ [Microsoft-Specific Search Engine](#) ▶ [University Search Feature](#)

Search Features

- ▶ [Cached Pages](#) ▶ [Date Restricted Search](#) ▶ [Definitions \(Google Glossary\)](#) ▶ [Diacritics searching \(terms with accents\)](#) ▶ [Dictionary Link](#) ▶ [File type conversion](#)
- ▶ [New Froogle home page](#) ▶ [Froogle Product Reviews](#) ▶ [Froogle Wish Lists](#) ▶ [Make Google Your Default Engine](#) ▶ [_ Operator](#) ▶ [Multilingual Google Definitions](#)
- ▶ [My Search History](#) ▶ [News Headlines](#) ▶ [Images in Google News search results](#) ▶ [Google News: Personalized News](#) ▶ [News Alerts](#) ▶ [Number range \(numrange\) command](#)
- ▶ [OR Operators](#) ▶ [Phonebook](#) ▶ [Personalized Web Search](#) ▶ [Phrase Search](#) ▶ [+ Operator](#) ▶ [Product Search Shortcut](#) ▶ [Google Q&A](#)
- ▶ [Raised search limit to 32 words](#) ▶ [Results Prefetching](#) ▶ [SafeSearch Filtering](#) ▶ [Satellite View from Google Maps](#) ▶ [Search by Location](#) ▶ [Search by Number](#)
- ▶ [Search stop words in phrases](#) ▶ [Similar Pages aka GoogleScout](#) ▶ [Site-Flavored Google Search Box](#) ▶ [Spelling Corrections](#) ▶ [Stock Quotes](#) ▶ [Live stock quotes](#)
- ▶ [Street Maps](#) ▶ [Google Suggest](#) ▶ [Google Suggest in Japanese](#) ▶ [~ Synonym Operator](#) ▶ [Translation](#) ▶ [Useful links from within a site](#) ▶ [User Interface \(UI\) redesign](#)
- ▶ [Web Alerts](#) ▶ [* Wildcard \(currently not working\)](#) ▶ [Word Variation \(Automatic Stemming\)](#)

Communicate, Show & Share

- ▶ [GMail](#) ▶ [Google Groups](#) ▶ [Google Groups 2](#) ▶ [Added mailing list support to Google Groups](#) ▶ [Language Tools](#) ▶ [Orkut \(Social Networking Service\)](#) ▶ [Picasa 2](#)
- ▶ [Google Talk](#) ▶ [Translation](#)

Mobile Services

- ▶ [Google Mobile - Local Search](#) ▶ [Google SMS](#)

For Website Hosts

- ▶ [AdWords](#) ▶ [Google Sitemaps](#)

Other Google Features

- ▶ [Google Answers](#) ▶ [Google Answers Searchable](#) ▶ [Browser Buttons](#) ▶ [Calculator](#) ▶ [\\$ Currency Conversion](#) ▶ [Deskbar](#) ▶ [Google Desktop 2](#) ▶ [Google Desktop Search](#)
- ▶ [Enterprise Desktop Search Tool](#) ▶ [Google Help: Cheat Sheet](#) ▶ [Home page tabs](#) ▶ [Google Labs](#) ▶ [Personalize Your Homepage](#) ▶ [Preferences](#) ▶ [Toolbar](#)
- ▶ [Toolbar for Firefox](#) ▶ [Toolbar 2.0 with a pop-up blocker](#) ▶ [Google Viewer](#) ▶ [Google Web Accelerator](#)

• http://www.googleguide.com/feature_history.html

• <http://de.wikipedia.org/wiki/Google-Dienstleistungen>

Wie arbeitet Google?

- **PageRank** untersucht Beziehungsgeflecht im Web;
- ein Link auf eine Seite ist wie eine „Stimme“ für diese Seite;
- Eine „Stimme“ von einer oft verlinkten Seite zählt mehr als die Stimme „unwichtigen“ privaten Seite;
- die Seite mit den meisten „Stimmen“ wird oben gelistet;
- Kein Mensch greift ein. Alles geschieht nach einem vorgegebenem Algorithmus;

Aber Achtung! „Popularität“, allein, ist aber noch kein Hinweis auf Qualität!

Richtig suchen

1. Wissen, was wir suchen
2. Die richtige Strategie wählen
3. Das richtige Werkzeug oder die richtigen Werkzeuge wählen

Suchdienste

- Suchmaschine

Beispiel: [http:// www.google.ch](http://www.google.ch)

siehe: <http://de.wikipedia.org/wiki/Google>

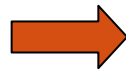
- Suchkatalog (Webverzeichnis)

Beispiel: <http://directory.google.com>

Suchstrategien

- **Thematische Suche**

Möglichst umfassende Sammlung von zuverlässigen Dokumenten zu einem gewissen Thema



Webverzeichnis

- **Gezielte Suche**

Finden einer spezifischen Information.



Automatische Suchmaschine

Problem: Treffermenge zu klein

- Anderen Suchbegriff wählen
- Alternative Schreibweisen für den Suchbegriff wählen
Beispiel: "*Djihad/Dschihad*" oder "*Beijing/Peking*"
- Erweiterte Suche bei Google „mit irgendeinem der Wörter“
- Andere Suchmaschine wählen
- (z.B.: Metasuchmaschine vgl. <http://us.ixquick.com/deu/>)

Problem: Treffermenge zu groß

- Kombination von mehreren Suchbegriffen („und“)
 - *"Weber" zu viele Ergebnisse, tragen Sie zusätzlich den Vornamen "Max" sowie ggf. das Schlagwort "Soziologie,"*
„Max Weber“ Soziologie
 - Oder: yesterday lyrics Beatles
- Phrasensuche
- erweiterte Abfragemöglichkeiten bei Suchmaschinen
 - Google: „Erweiterte Suche“
 - Term-minator: „Vertikale Suche“
(<http://www.term-minator.eu/de.htm>)

Suchmöglichkeiten

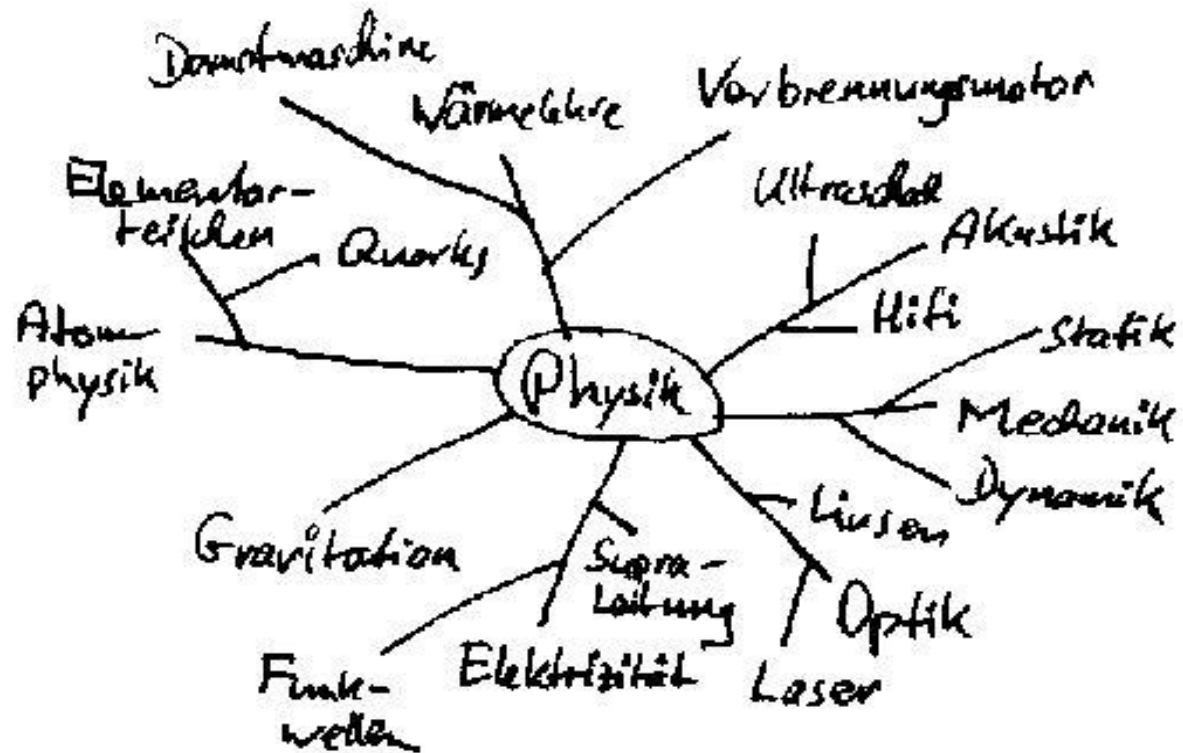
automatische Suchmaschine

- Einfache Suche
- Erweiterte Suche: http://www.google.de/advanced_search?
- Benutzung der Suchoperatoren:
<http://www.term-minator.it/D-Google-Operatoren.htm>
- Anordnung der Keywords in mind maps
- “Suchoptimierte Suche”:
<http://www.term-minator.eu/de.htm>

Webverzeichnis

- Suche in den Kategorien

Mind map: Begriffe assoziieren



Alternativen zu Google

- [http:// www.yahoo.de](http://www.yahoo.de)
- <http://ch.msn.com/>
- <http://ch.altavista.com/>
- Kombinierte Suche in verschiedenen Rubriken: <http://www.a9.com>
- Clustering von Suchergebnissen: <http://www.clusty.com>
- [http:// www.ask.com](http://www.ask.com)
- [http:// www.answers.com](http://www.answers.com)
- Verzeichnis von Suchmaschinen
<http://www.suchlexikon.de/>
- Grafische Visualisierung der Treffer und deren Verbindungen
Kartoo: <http://www.kartoo.com/>

Evaluierung einer Website

- Wer ist die SenderIn oder AutorIn?
- Informationen in der URL:Name, Domäne (was:.com, .edu, .org,...; woher:.ch, .de,,,fr,.at,,it,,.uk);
- Ist er oder sie bekannt und angesehen auf seinem/ihrer Gebiet?
- Ist es möglich sich mit ihm/ihr in Verbindung zu setzen?
- Wird die Site regelmäßig betreut?

- Vgl.: <http://www.ssm-site.ch/maz/quellenpruefung.html>
- Check mit: http://www.term-minator.eu/de_check.html

Literatur

Holger Bleich: *Bosse der Fasern. Die Infrastruktur des Internet*. In: c't 7/2005, S. 88-93 (21. März 2005)

Tim Berners-Lee: *Information Management: A Proposal*, 1989

Tara Calishain / Rael Dornfest: *Google Hacks. 100 Insider-Tricks & Tools*, O'Reilly 2003.

Sebastian Erlhofer: *Suchmaschinen-Optimierung für Webentwickler. Grundlagen, Ranking optimieren, Tipps und Tricks*, Galileo Press 2006.

Gina Maneri / Hellmut Riediger: *Internet nel lavoro editoriale. Risorse, strumenti, strategie per redattori, traduttori e per chi lavora con il testo*, Milano: Editrice Bibliografica 2006; Online-Auszug
<http://www.edigeo.it/Sitoteca/sitoteca.php?lingua=&sez=risorse>.

Christoph Meinel, Harald Sack: *WWW – Kommunikation, Internetworking, Web-Technologien*, Springer-Verlag, Berlin, Heidelberg, New York, 2004.

Hellmut Riediger: *Internet per mediatori linguistici*, Winterthur: ZHW 2006; Online-Kurs (italienisch): <http://www.term-minator.it/corso>