

Effektiv Suchen im Internet

Suchstrategien, Methoden, Quellen

Zusammengestellt während des eigenen Lernens und den dabei auftauchenden Fragen von Jürgen Schramm (Schrammii@web.de).

Nach einem empfehlenswerten Buch von Ulrich Babiak, O'Reilly Verlag, 1997.

Grundsätzliches

Groß- und Kleinschreibung beachten!

Am Ende einer URL immer ein /

Bei Problemen: Ist die Verbindung zustande gekommen?

Abhilfe: In der Verzeichnisstruktur klettern!

Mit Suchmaschinen nachhelfen!

Alle hier erwähnten URL's findet ihr in meinem Bookmarkverzeichnis!

Was findet man eigentlich im Internet?

Kunstgegenstände <http://mistral.culture.fr/cgi-bin/mistral/joconde>

David Copperfield <http://www.snafu.de/~tilman/zauber/index.html>

Gesetze <http://sunsite.informatik.rwth-aachen.de/Knowledge/germlaws/>

Warentests http://www.dbi-berlin.de/dbi_dbf/dbi_dbf.htm

Musik und CD's <http://www.cdconnection.com/>

Für eine effektive Suche im Internet ist das Verstehen Was, Wo, Wie, von Wem und aus Welchem Grund im Internet steht sehr hilfreich!

Warum ist das Suchen so kompliziert?

Im Internet fehlt die Organisation, die Strukturierung. Dann die Beliebigkeit und die Dynamik des Netzes.

Mehr als WWW - Wichtige Internetdienste

Gopher

Der textbasierte Vorläufer des WWW. Wenn überhaupt, dann nur noch via WWW genutzt.

WWW

Andere Dienste wie Abfragen usw. sind auch meist via WWW möglich. Gateways übernehmen die Wandlung.

eMail und Mailinglisten

Grundlagen mit Beispielen: <http://fr-koeln.de/fbi/hilfe/manual/biman54.html>

die größte Liste mit 66.000 Mailinglisten: <http://www.liszt.com/>

Liste, auch eMail: <http://www.n2h2.com/KOVACS/>

Tile.Net mit allen Listen, die die Software *listserv* (Crosspoint!) verwenden:

<http://tile.net/lists/>

List öffentlicher Listen:

http://www.im.net/cgi-bin/local_mailinglists

Suchen einer Mailingliste in Suchmaschinen mit dem Stichwort *Mailingliste* oder *Mailing-Liste*.

Newsgruppen

Sind unterteilt in sogenannte Hirarchien:

alt	Alternative
biz	comp misc
rec	Hobby und Freizeit
sci	Wissenschaftliche Foren
soc	Kultur und Soziales
de	deutschsprachige Gruppen
cl	
maus	
z-netz	Z-Netz-Forum

Übersichten gibt's unter <http://www.w3.org/pub/DataSources/News/Groups/Overview.html>.
Deutschsprachige Gruppen der de-Hirarchie unter <http://search.dejanews.com/bg.xp?level=de>.

FTP

Der Schlüssel zu Software- und Dateiarchiven! Erfolg gibt's nur mit genauen Quellangaben.

FAQ's gibt's, nach Gruppen geordnet, auf dem FTP-Server [ftp.uni-paderborn.de](ftp://ftp.uni-paderborn.de) im Verzeichnis `/pub/FAQ`.

Telnet

Auch heute noch eine wichtige Sache, die direkte Verbindung zu einem Rechner. Z.B. zur Abfrage eines Bibliothekskataloges usw.

WAIS und Z39.50

Die US-ANSI-Norm Z39.50 und ihr Ableger WAIS dienen der Bereitstellung von Datenbanken über Netzwerke. Es sind praktisch Client-Server-Anwendungen, die eine Client-Software auf deiner Seite benötigen. Einen Überblick gibt es via WWW-Gateways wie z.B: http://munin.ub2.lu.se/auto_new/UDC.html. Dort gibt es einen Übergang zu WAIS und eine gute Übersicht über die weltweit verfügbaren WAIS-Datenbanken.

Über <http://europagate.dtv.dk/> und <http://lcweb.loc.gov/z3950/agency/> gibt es experimentelle Übergänge zur Suche in mehreren Datenbanken, u.A. Bibliotheken.

Schneller Sucherfolg mit Basistechniken

2 grundlegend unterschiedliche Techniken:

Suchmaschinen

massive Unterstützung von Maschinen

Suche erfolgt im Volltext

vor allem kommerziell für die Anbieter interessant.

Daneben viele kleinere und spezielle Suchmaschinen

die wichtigsten Suchmaschinen:

AltaVista, Excite, HotBot, Infoseek Guide, Lycos, Opentext, Infoseek Ultra und *Webcrawler*.

Trotz eines berechtigten Vollständigkeitsanspruches verlasse dich nie auf eine Suchmaschine allein! Neben dem Suchbegriff kann meistens das Suchgebiet (WWW oder Newsgruppen) angegeben werden. Kleinschreibung!

thematische Verzeichnisse

Hier wird im Gegensatz zu Suchmaschinen noch mit Menschenhand gearbeitet. Eine Suche in nach themensortierten Übersichten (Katalogen) ähnelt sehr stark den bekannten Gelben Seiten. Ihr Aufbau ist bekanntlich hierarchisch.

Solche Verzeichnisse sind sehr gut geeignet, wenn gar keine genaue Anforderung besteht, sondern mehr eine Übersicht über ein Gebiet und deren Angebot im Internet gewünscht ist. Querverweise dienen dazu, auch in einer untergeordneten Kategorie noch den (richtigen) Weg zu finden.

die wichtigsten thematischen Verzeichnisse:

Yahoo!, *WWW Virtual Library*, *Tradeware Galaxy*, *Yahoo* Deutschland, *Web.de* und *DINO*.

Yahoo ist der größte und populärste Anbieter, die *WWW Virtual Library* dagegen eine nicht kommerzielle Zusammenstellung mit wissenschaftlichem Schwerpunkt.

Yahoo.com und *Yahoo.de* sind miteinander verbunden. Eine Suche im deutschen Yahoo und ein Querverweis auf *Yahoo.com* funktionieren automatisch.

Die *WWW Virtual Library* ist das älteste Verzeichnis, entwickelt von den Schöpfern des WWW in Genf. Hinter dem Verzeichnis steckt kein Rechenzentrum, sondern es ist verteilt über die ganze Welt! Der Einstieg erfolgt über <http://www.w3.org/vl/>. Es kann zwischen einer Gesamtdarstellung oder einer Liste nach Kategorien gewählt werden. Die URL <http://www.rz.uni-karlsruhe.de/Outerspace/VirtualLibrary/> bietet den Einstieg zu deutschen Datenquellen des wissenschaftlich-akademischen Bereiches.

Weitere deutsche Anbieter sind *DINO* und *Web.de*. Sie sind zwar klein, aber für Suchen nach deutschen Quellen flächendeckender.

Du weißt nicht, in welchem Katalog suchen? Dann hilft Dir eine Stichwortsuche im thematischen Verzeichnis weiter. Unterschied zur Suche in Suchmaschinen beachten!

Eine Stichwortsuche in *AltaVista* oder *Excite* durchsucht die gesamten Internetdokumente. Eine Suche in *Yahoo* oder *Web.de* durchsucht hingegen nur die Einträge und Beschreibungen in *Yahoo* oder *Web.de* selbst.

Wie Suchmaschinen funktionieren und welche Grenzen sie haben

Gewöhne dich an den Gedanken, niemals alle gefundenen Verweise lesen zu können! Du mußt in der Lage sein, ab einem gewissen Schlußstrich alles Ungelesene zu verwerfen.

Die Trefferliste ordnet nach Güte (Ranking). Sie arbeitet nach der Anzahl der gefundenen Wörter, deren Position, dem Abstand und ihrer Häufigkeit.

Meist findet ihr beim Zitat einen Schalter wie *More like this*, der weitere ähnliche Dokumente sucht (Relevanz-Feedback). Die ausgefeiltesten Verfahren hat (oder hatte?) *Excite*.

Ist ein Server den Suchmaschinen nicht bekannt, gibt es auch nichts zu finden. Geänderte Dokumente sind oft erst mehrere Wochen später recherchierbar. Bereiche können auch ausgeschlossen werden, ebenso werden geschützte Daten nicht erfaßt. Bereiche auf Servern, zu denen nur angemeldete Nutzer Zugang haben, werden ebenfalls nicht erfaßt. **Inhalte** von via WWW erreichbaren Datenbanken sind ebenso wenig erfaßt. Deren eigene Suchkataloge können jedoch ausfindig gemacht werden.

Direkte Suche nach URL's können nur über *HotBot*, *AltaVista* oder *Infoseek Ultra* erfolgen.

Es gibt keine „beste“ Suchmaschine. D.h. die Beste ist, die ihre Suche beantworten kann!

Die Sprache ist eigentlich egal, wird doch auf einen deutschen Suchbegriff meist auch nur ein deutsches Dokument gefunden. Von einigen Suchmaschinen entstehen deutsche Ableger (z.B. *Lycos.de*). Rein deutsche Suchmaschinen sind z.B. *Flipper*, *Hotlist*, *Kolibri*, *Crawler* und *Aladin*. In internationalen Suchmaschinen kann die Suche auf die Domain *.de* beschränkt werden.

Newsgruppen

Suchmaschinen für Newsgruppen weisen eine andere Indizierung auf. Suchmaschinen enthalten meist Alles aus den letzten vier Wochen, Newsserver dagegen evtl. kürzer! Neuere Anbieter wie *DejaNews* (<http://www.dejanews.com/>) planen die Archivierung aller News. **Denkt daran, daß so vielleicht noch eure Kinder, potentielle Arbeitgeber oder Kunden eure Beiträge lesen können!**

Agenten

Einen ersten Testagenten findest du unter <http://bf.cstar.ac.com/bf/>

Das Suchprogramm *Echo Search* gibt es als Demoprogramm unter <http://bushido.iconovex.com/ECHO/ECHO.HTM>. Es arbeitet vom eigenen Rechner aus, kennt mehrere Suchmaschinen und ist äußerst bequem!

Besprechungsdienste

Gewichtete Informationen werden heute immer wichtiger. Es gibt noch wenige, auch *Besprechungsdienste* genannte Dienste. Sie bestehen aus Besprechungen und Beurteilungen und stellen praktisch eine Bestenliste dar. Sie sind z.Z. noch kostenlos und durch Werbung finanziert oder von Medienkonzernen unterhalten.

Webtip (<http://www.webtip.de/>)

Ein Besprechungsdienst der Zeitschrift *DOS*. Noch sehr klein und übersichtlich.

Lycos Top 5% (<http://www.pointcom.com/categories/>)

Auch *Point Reviews* genannt. Lycos bietet damit eine Suchmaschine, ein Themenregister und diesen Infodienst an. Leider oft verwechselt.

Magellan (<http://magellan.mckinley.com/>)

Der größte Dienst mit dem Anspruch, möglichst viele Dokumente zu erfassen.

Argus Clearinghouse (<http://www.clearinghouse.net/>)

Eine besondere und wichtige Form von Suchhilfe! Es behandelt alle Internet-Dienste gleich, also neben WWW auch Gopher, News und Mailinglisten. Es entstand als Studentenarbeit, die sich dann schnell verselbständigte. Hier findet ihr zu vielen populären und wissenschaftlichen Themen umfangreiche Zusammenstellungen von Internetquellen, sozusagen vorgefertigte Komplettrecherchen!

Collaborative Filtering

Falsch übersetzt heißt das Empfehlungs- oder Bewertungssystem. Solch ein System macht dir Vorschläge, die auf den Vorlieben anderer User basieren. Sie funktionieren erfahrungsgemäß erstaunlich gut!

Bevor aber Wertungen abgegeben werden, müssen User erst einmal einen Teil ihrer Privatsphäre aufgeben. Das ist ein sehr heikles Thema. Das System muß dich kennen, um dir wirklich geeignete Vorschläge machen zu können. In der Regel muß du also etliche eigene Beurteilungen abgeben, bevor ein treffender Vorschlag zurück kommt.

Es gibt bereits funktionierende Systeme zum Themenkreis Film, Chat (Diskussionsräume), Bücher, Spiele, Warensysteme, News aus dem UseNet sowie allgemeine Informationen und auch zu Userbereichen! Darin könnt ihr vielleicht schnell Euch interessierende andere User kennenlernen.

Die Systeme (URL's siehe Bookmarksammlung) unterscheiden sich dabei in ihren Schwerpunkten. Es ist auf alle Fälle mal einen Test wert und die Überwachung und Datenspeicherung ist (noch) nicht so perfekt, daß sich euer Userprofil nicht bald wieder „ausgewachsen“ hat.

Suchhilfen für Suchhilfen

Es gibt Listen von Suchhilfen in den thematischen Verzeichnissen aber auch sogenannte Meta-Suchhilfen. Sie verweisen als Ergebnis deiner Nachfrage auf Fundstellen in anderen Suchmaschinen, Verzeichnissen usw. oder übernehmen gar gleich deren Durchführung selbst.

Dabei werden teilweise standardisierte Formulare benutzt. Oft ist es sinnvoll, noch direkt bei den erkannten Suchquellen nachzusuchen.

Die bekanntesten sind *Internet Sleuth* und *Search.com*, die älteste ist wohl *CUSI*. Vorteil dieser Suchhilfen ist, daß damit auch Hinweise auf Datenbanken gefunden werden, die man gar nicht kannte! Versucht doch mal eine Volltextdatenbank für die Bibel zu finden.

Stichwortsuchen

Stichwortsuchen erfolgen meist in Volltextdatenbanken, also wenig bis gar nicht strukturierten Datenbeständen. Recherchen in besser strukturierten Datenbanken haben meist ihre eigene Suchsprache.

Es ist empfehlenswert, sich auf das bescheidene sprachliche Niveau der Rechner herunterzugeben. Zu unterscheiden sind die freie Eingabe mit spezieller Syntax und die menüunterstützten Eingaben. Manchmal ist auch deren Wahl möglich. Menüsuchen sind einfacher, aber auf die vorgegebene Struktur beschränkt.

Tip: suche dir eine Suchmaschine aus, studiere deren Suchsyntax genau und du wirst mehr Erfolg haben als bei oberflächlichen Suchen in diversen Maschinen. Fang mit *AltaVista* an und teste nach einige Zeit eine Andere.

Maskierung (Trunkierung)

Hierbei wird die Endung eines Suchwortes durch ein * ersetzt, damit deren unterschiedliche Formen gefunden werden.

Siehe *AltaVista*, *Opentext*, *Lycos* (automatisch, ohne *), *DINO*..

Substringsuche

Diese Suchform muß meist irgendwo angekreuzt werden. Sie entspricht dann der beidseitigen Trunkierung. Meist ist nur Eine dieser Formen erlaubt.

Siehe *Yahoo* (kann abgeschaltet werden), *Web.de*.

Manche Dialekte lassen auch die beidseitige Form *Suchwort* zu (*DINO*) oder gar eine Trunkierung in der Mitte eines Suchwortes (*AltaVista*, *DINO*).

Kombination mehrerer Suchbegriffe

Unschärfe Suche

Hier werden einfach alle infrage kommenden Wörter (durch Leerzeichen getrennt) hintereinander geschrieben. Das Ergebnis ist also nie 100%, die Fundstellen werden aber nach Güte sortiert angezeigt. Interessant sind meist also nur die ersten Fundstellen und zwar umso mehr, als du alle(!) infrage kommenden Begriffe auch wirklich genannt hast. Ist ein gesuchtes Dokument nicht unter den ersten 10 bis 20 Fundstellen muß die Suche (variiert) wiederholt werden. Für solche Suchen sehr gut geeignet (war) *Excite*, mit seinen für diese Suchweise optimierten Algorithmen.

Wichtig ist die Suche so einschlägig wie möglich zu gestalten! Und das erfolgt durch Eingeben möglichst vieler(!) zutreffender Suchbegriffe. Ich hab auch Anfangs falsch gedacht :-). Ohne Einschränkung kann gesucht werden in *Excite*, *Infoseek* und *AltaVista* (Simple Query). Die Reihenfolge der Begriffe und deren Sprache spielt keine Rolle!

Eingeschränkt oder gar nicht nutzbar ist diese Suchform bei *Opentext*, *AltaVista* (Advanced Query), *Lycos* sowie allen thematischen Verzeichnissen. Bei *Hotbot* den Wahlschalter (any of these words) beachten und ausprobieren. Beachte die unterschiedliche Form bei *AltaVista*.

Trennscharfe Suche

Hierbei gibt es nur trennscharfe und eindeutige Ergebnisse. Die Boole'schen Operatoren AND, OR und NOT erlauben die volle Kontrolle über das Ergebnis.

Egal wie die Suchbegriffe definiert werden, es gibt eine genau berechenbare Menge an Fundstellen.

Der **OR-Operator** faßt die Treffermengen mehrerer Suchbegriffe zusammen. Es bedeutet nicht Entweder-Oder, sondern einer der beiden oder auch beide Begriffe! In der Praxis werden somit oft unterschiedliche Schreibweisen berücksichtigt. Z.B. Hose OR Beinkleid, oder book OR Buch.

Mit OR wächst also die Treffermenge, ohne daß sie qualitativ besser wird. Aber **Vorsicht:** in vielen Datenbanken ist OR die standardmäßige Verknüpfung bei mehreren Suchbegriffen!

Der **AND-Operator** schränkt die Treffermenge ein. Das Ergebnis wird also besser und erlaubt auch bei mehreren Suchbegriffen ein zielgerichtetes Suchen. Die Treffermenge wird genauer. **Achtung:** Schreibfehler wirken sich hier besonders stark aus!

Der **NOT-Operator** schließt unerwünschte Aspekte aus. Gerade in Suchmaschinen tauchen oft Treffer auf, die mit dem eigentlich gesuchten Thema nichts zu tun haben.

AltaVista Simple Query, *Infoseek* und *Excite* benutzen das +Zeichen vor einem Begriff für ein AND, entsprechend ein -Zeichen für NOT. Ohne Zusatzzeichen ist es eine unscharfe Suche, in etwa ein OR!

Bei *Excite* und *Webcrawler* stehen die Operatoren einfach zwischen den Begriffen, ebenso bei *AltaVista* Advanced Query, anstelle von NOT muß es aber AND NOT heißen.

DINO verknüpft Begriffe automatisch mit AND, OR und NOT sind möglich.

Yahoo und *Opentext* verknüpfen ebenfalls automatisch mit AND.

Nicht oder nur eingeschränkt verfügbar sind diese Funktionen bei *Web.de*.

Die **Abstandsoperatoren NEAR und ADJ** geben einen Abstand für die Suchbegriffe in Trefferdokumenten an. NEAR ist präziser als AND und steht für „nahe bei“. Der Abstand beträgt meist 25 bis 100 Wörter und kann manchmal (*Webcrawler*) auch vorgegeben werden.

Diese Funktion ist nicht verfügbar bei *Excite*, *Lycos*, *Hotbot* und allen thematischen Verzeichnissen.

Phrasen und Bezeichnungen können mit den vorgenannten Funktionen nicht gesucht werden. Feststehende Begriffe wie Hewlett Packard, Afri Cola oder eine bestimmte Webseite (die du noch irgendwie im Gedächtnis hast) werden so zielsicher gefunden.

AltaVista, *Webcrawler*, *Infoseek* und *Excite* brauchen einfach die Begriffe in „Anführungszeichen“. In *Web.de* werden Begriffe automatisch so gesucht. Nicht oder nur eingeschränkt verfügbar ist die Phrasensuche bei *Lycos*, *Yahoo* und *DINO*.

Umlaute, Sonderzeichen und Kurzworte

Suchbegriffe mit Umlauten sind ein Problem. Keine Unterstützung ist von *Excite* und *Lycos* (USA) bekannt. Sonderzeichen wie Bindestriche, Punkte oder Pluszeichen sind nicht suchbar. Die Suchmaschinen ignorieren sie, es sei denn, sie werden als Phrasen angegeben! Z.B. „CD-ROM“, „Linux 2.0.13“ oder „Intel P166“.

Auch Kurzworte mit weniger als drei Buchstaben werden im allgemeinen ignoriert. Eine Suche nach Windows NT muß daher als Phrasensuche erfolgen.

Einschränkungen

Bei „normalen“ Datenbanken üblich, im Internet noch kaum, ist die Einschränkung einer Suche auf bestimmte Dokumentteile oder Felder.

Möglich sind bei WWW-Dokumenten Titel, URL, Hostname, Dateityp, Erstellungsdatum, Hypertextlink oder Zwischenüberschrift. Bei Newsgruppen sind Einschränkungen auf Newsgruppe, Autorennamen, eMail-Adresse und Betreffzeile möglich.

Dies ermöglicht sehr spezielle Suchen! So enthält ja HTML einen <TITLE>-Tag zum Suchen allein in diesen Titeln, oder du beschränkst Suchanfragen auf bestimmte Datumzeiträume oder Dateitypen.

Suchanfragen mit Klammern

...

Zusammenfassung

...

Recherchestrategien

...

Beispielrecherchen

...

Software finden

...

Personen finden

...

Homepage, eMail-Adressen

...

Wie man sich auf dem Laufenden hält

...

Beurteilung von Suchergebnissen

...

Kommerzielle Anbieter

...

Zukunft

...

Stand Sommer 1999