

Einführung Internetrecherche

Was geht und was nicht?

Was ist im Netz durchsuchbar?

- Das World Wide Web, d.h. über HTTP erreichbare Seiten
- News-Archive
- FTP-Archive
- Mail-Archive

Was ist in den Ergebnissen von Suchmaschinen und Katalogen nicht enthalten?

- relativ neue Seiten
(sehr unterschiedliche Aktualisierungszyklen)
- Seiten, die auf tieferen Ebenen einer Website liegen
- Inhalte von Datenbanken (z.B. Pressearchive)
- Seiten hinter Firewalls (hausinterne Informationen)

Kataloge und Suchmaschinen

Grundsätzlich wird zwischen Katalogen und Suchmaschinen (Spidern, Robots) unterschieden.

Kataloge (Yahoo, web.de, dmoz.org)

- Mit echten Menschen
- Redaktionell bearbeitete und bewertete Informationen
- Seiten sind Inhaltlich/thematisch geordnet
- **Anwendung:** Überblick über ein Thema bekommen

Suchmaschinen (Google, Altavista, Fireball, Lycos)

- Automatische Informationssammlung
- Die Mathematik ist entscheidend, nicht der Inhalt
- **Anwendung:** Suche nach konkreten Informationen zu einem Sachverhalt

Kataloge: Yahoo

Hierarchisch aufgebaute und redaktionell bearbeitete Linkliste

- Suchfunktion für die Einträge im Katalog

Beispiele

- [Yahoo!](#) > [Gesellschaft und Politik](#) > [Kriminalität](#) > [Terrorismus](#) > [Terroranschlag auf New York und Washington, 11. September 2001](#) > Falschmeldungen
- [Yahoo!](#) > [Gesellschaft und Politik](#) > [Behörden](#) > Nachrichtendienste und Geheimdienste
- [Yahoo!](#) > [Nachrichten und Medien](#) > [Journalismus](#) > Recherche
- http://de.dir.yahoo.com/Internet_und_Kommunikation/Clubs_und_Diskussionsforen/

Suchmaschinen:

Wie funktionieren Spider, Robots und Co?

- Google - Suchen auf
Zweimilliardenvierhundertneunundsechzigmillionenneunhundertvierzigtausendsechshundertfünfundachtzig Web-Seiten in 0,48 Sekunden.
- Voraussetzung: große Rechnerfarm, Gigabitweise RAM, ein richtig dickes Kabel zum Internet und ein raffinierter Algorithmus

Vier Schritte

- Informationen sammeln
- Daten indizieren
- Suchanfragen bearbeiten
- Ergebnisse bewerten

Suchmaschinen: Sammeln!

Spider hangeln sich von Link zu Link durchs Netz.

Seiten ohne Link oder Anmeldung bei Suchmaschinen werden nicht gefunden!

- Suchtiefe bis zu drei Ebenen
- Zyklus: durchschnittlich vier Wochen

Suchmaschinen: Indizieren!

Die Indizierungssoftware strukturiert die übermittelten Daten und macht sie durchsuchbar.

- Die Seiten werden in einer Datenbank gespeichert.
- Bei einer Suchanfrage wird die Datenbank und nicht das Internet durchsucht.

Suchanfrage bearbeiten!

Die Suchanfrage besteht gewöhnlich aus mehreren Suchbegriffen die mit und/oder /nicht verknüpft sind.

- Die Datenbank wird nach diesen Kriterien durchsucht und gibt eine Ergebnisliste zurück.

Bewertung der Ergebnisse?

Nach welchen Kriterien werden die Ergebnisse bewertet?

Eigenschaften des Dokuments

- Werden alle Suchbegriffe auf der Seite gefunden?
- Wo stehen die Suchbegriffe (Titel, Metatext, Textanfang, etc)?
- Wie oft werden die Suchbegriffe auf der Seite gefunden?

Plazierung von Treffern durch Bezahlung

Listung in Katalogen

- Z.B. Suchbegriff entspricht dem Kategorienamen in dem die Seite eingeordnet ist

Nutzerverhalten

- Wie oft wird ein Ergebnis angeklickt?
- **Page Rank:** wie oft ist eine Seite verlinkt?

Bewertung der Ergebnisse?

Google zu Page Rank:

- PageRank verlässt sich auf die einzigartige demokratische Natur des World Wide Webs, indem es die weitverzweigte Link-Struktur als einen Indikator für die individuelle Einschätzung der Qualität einer Seite nimmt. Der Kern ist dabei, dass Google einen Link von Seite A zu Seite B als ein "Votum" von Seite A für Seite B interpretiert. Aber Google sieht sich mehr als nur das Ausmaß der Zustimmung oder der Links auf einer Seite an; Google analysiert ebenfalls die Seite, die das Votum abgegeben hat. Das Votum von einer Seite, die selber "wichtig" ist, zählt mehr und hilft, andere Seiten "wichtig" zu machen.

Zusammenfassung der Kriterien

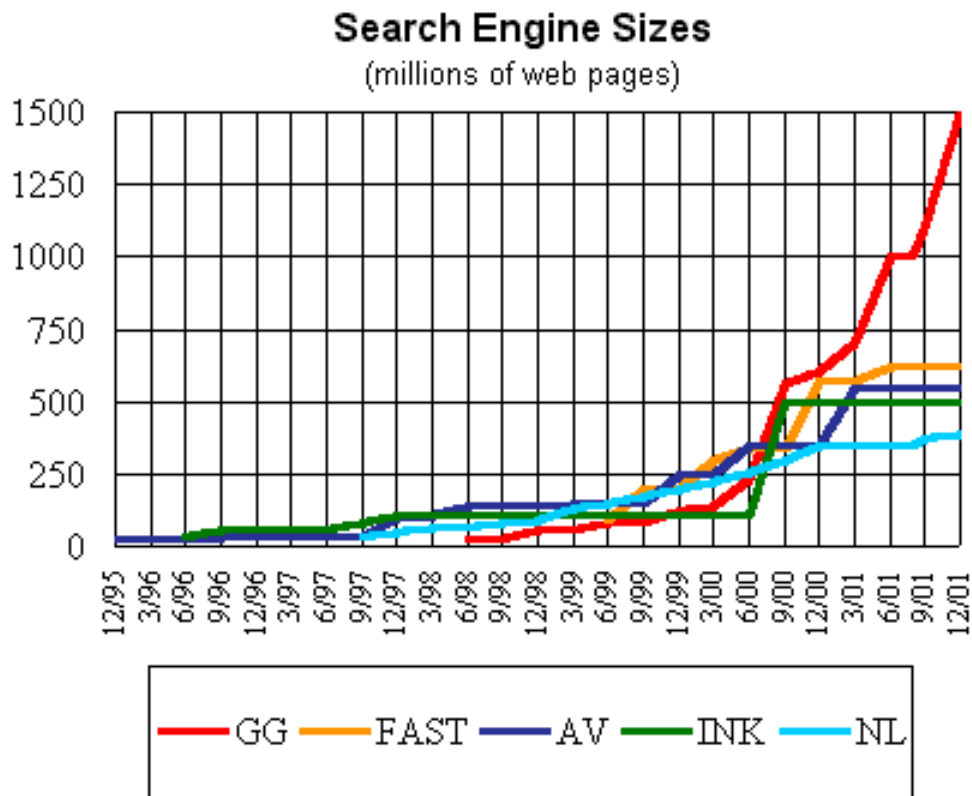
Bewertung der Ergebnisse nach:

- Position und Häufigkeit des Suchbegriffs im Dokument
- Vorkommen in einem der Kataloge
- Klickrate
- Page Rank
- Bezahlung

Verschiedene Suchmaschinen, verschiedene Kriterien, verschiedene Ergebnisse

- Die genaue Funktionsweise und Gewichtung ist Betriebsgeheimnis

Entwicklung



GG: Google

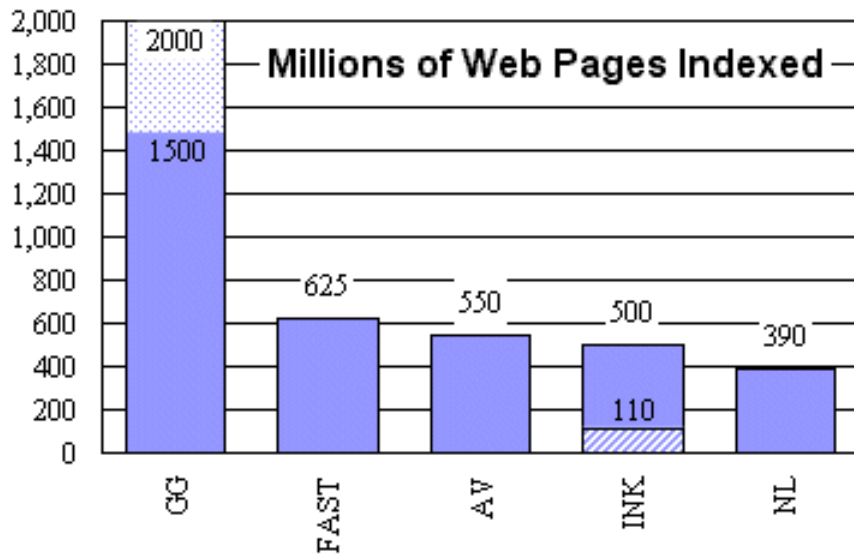
Fast: Lycos, Hot Bot, Excite

AV: Altavista, web.de (int.)

Inktomi: Hot Bot (bis Sep)

NL: Northern Light

Stand 12.2001



GG: Google

Fast: Lycos, Hot Bot, Excite

AV: Altavista, web.de (int.)

Inktomi: Hot Bot (bis Sep)

NL: Northern Light

Quelle: www.searchenginewatch.com

Wie suchen?

Der Einstieg

Kataloge:

- Allgemeinen Überblick gewinnen

Suchmaschinen:

- Suche nach konkreten Informationen zu einem Sachverhalt

Die Suche beginnt im Kopf

Gesuchtes Thema genau einkreisen, Stichwörter machen

Sprache, in der das Zielobjekt abgefasst sein darf?

Synonyme und andere Begriffe aus der Thematik sammeln

- *Niveau* (Klimaforschung statt Wetter),
- *Fachausdruck* (Meteorologie statt Wetter)
- *Präzisierung* (Klimaerwärmung statt Klima)

Schreibweisen der Suchbegriffe überdenken

- Verben und Substantive, Singular, Plural
- Google unterstützt keine Trunkierung: Die Suche nach „Auto“ ergibt keine Ergebnisse wenn nur der String „Autos“ auf der Seite vorkommt.

Was durchsuchen?

HTTP

- „normale“ Internetseiten

Bilder

- Text in der Nähe des Bildes wird zur Bewertung herangezogen

Newsgroups

- Google: Groups,

-

Programme, Software

- download oder herunterladen

Was wird den so gesucht?

Die häufigsten Suchanfragen bei Google

- <http://www.google.de/press/zeitgeist.html>

Live bei Fireball

- <http://www.fireball.de/>
LiveSuche anklicken

Suchmaschinen: Allgemein

Zu Beginn: Viele Begriffe, genaueres Ergebnis

- Beispiel: Presse, Medien, Gerüchte, Falschmeldungen, Internet,
- In welchem Zusammenhang stehen Themen und Begriffen

Phrasensuche

- „Was ist...“,
- „Wie funktioniert“
- „Vorname Name“

Boolsche Operatoren

- und/oder/nicht

Mehrere Suchmaschinen benutzen!!

Suchmaschinen: Boolsche Funktionen

und = Der Begriff muss vorkommen

- Google: Vorgabe (außer Stoppwörter)
- AV: +Gerüchte +Falschmeldungen
- Fireball Vorgabe

oder = Der Begriff kann vorkommen

- Google OR
- AV: Vorgabe
- Fireball OR

nicht = Der Begriff darf nicht vorkommen

- Google - (Minuszeichen)
- AV: - (Minuszeichen)
- Fireball - (Minuszeichen)

Suchmaschinen: Bediensyntax

Klammern: definieren eine Unterbedingung

- Auto +(rot OR grün) –VW

Site: Listet nur die Ergebnisse aus der Domain

- Site:www.spiegel.de Gerüchte Internet

Link: gibt die Seiten aus die einen Link auf die Adresse haben

- Link:www.heise.de

Suchmaschinen: Kein Treffer

Tippfehler? Kleinschreibung!

Suchbegriffe umformulieren (Singular/Plural), Umwege

Platzhalter benutzen

- Philo* findet Philosophie, Philologie, Philanthrop usw.

Zahl der Suchbegriffe reduzieren

Andere Suchmaschinen bzw. Metasucher nutzen

- <http://www.metager.de>

Suchmaschinen: Zu viele Treffer

Zusätzliche Suchbegriffe

Phrasensuche

Ausschließen von Themen, Bereichen

Suchmaschinen: Error 404

Diese Seite existiert nicht mehr

- verschoben, gelöscht????

Verzeichnisstruktur hocharbeiten

- Die Site behandelt u.U. das gesuchte Thema, nur die Seite ist nicht mehr vorhanden

Google: Im Archiv sind Seiten gespeichert die u.U. nicht mehr online sind

Suchmaschinen: Spezialisten

Internet

- Denic
- Google > Groups für die Suche im Usenet

Seriennummern

- Astalavista, warez.com

Mailadressen

- Mesa: <http://mesa.rrzn.uni-hannover.de/>
Es gibt kein Gesamtverzeichnis aller E-Mail Adressen (zum Glück)

Bilder im Netz

- Zu Layout und Präsentationszwecken
- <http://www.corbis.com/>, <http://www.gettyimages.com/>, <http://www.fotofinder.net>

Suchmaschinen: Spezialisten

Bücher und Zeitschriften:

- Paperball.de
- [Karlsruher Virtueller Katalog](#)
75 Millionen Buchtitel aus Bibliotheken und Buchhandel.
- Paperazzi.de
88 deutschsprachige Publikationen aus den Bereichen Nachrichten, Sport, Lifestyle, Business und Medien.
Auch ältere Artikel suchbar. Mit E-Mail-Benachrichtigungsdienst auf Stichworte.

Suchmaschinen: Spezialisten

Wissenschaft, Uni

- [Düsseldorfer Virtuelle Bibliothek](#)
Gut gepflegte Liste, besonders gut bei Geisteswissenschaften.
- [The WWW Virtual Library](#)
Universale Themenliste, die weltweit verteilt gepflegt wird. Je nach Gebiet unterschiedliche Qualität.
- <http://www.wissen.de>

Links zum Suchen

- <http://www.suchfibel.de>
- <http://www.suchmaschinen.de/>
- <http://www.suchmaschinentricks.de/>
- <http://www.ubka.uni-karlsruhe.de/suchmaschinen/index.html>

Newsgroups und Foren

Newsgroups (Usenet)

- Eigener Dienst
- Diskussionsforen des Internets
- Nach verschiedenen Gebieten hierarchisch geordnet

Foren

- Per HTTP erreichbare „Schwarze Bretter“

Foren und Open Postings

Foren: Beispiele

- http://de.dir.yahoo.com/Internet_und_Kommunikation/Clubs_und_Diskussionsforen/
- <http://parsimony.net/>

Open Posting:

- www.Indymedia.de

Newsgroups: Einrichten in Outlook

Genaue Anleitung: <http://www.newserverguide.de/>

Newsgroups: Themen

Welche Newsgroup für welches Thema eignet:

- news.groups.questions oder de.newusers.questions

Überblick

- comp.* und gnu.* für Computer und EDV,
news.* für Usenet News,
sci.* für Naturwissenschaften und Technik,
soc.* und talk.* für Geisteswissenschaften und persönliche Kontakte,
rec.* für Unterhaltung (recreation),
misc.* und alt.* für sonstige Themen.

Newsgroups: Regeln

Die Netiquette

- Vergessen Sie niemals, dass auf der anderen Seite ein Mensch sitzt!
- Erst lesen, dann denken. Noch einmal lesen, noch einmal denken. Und _dann_ erst posten!
- Teilen Sie etwas Neues mit!
- Ihre Artikel sprechen für Sie - Seien Sie stolz auf sie!
- Nehmen Sie sich Zeit, wenn Sie einen Artikel schreiben!
- Vernachlässigen Sie nicht die Aufmachung Ihres Artikels!
- Achten Sie auf die "Subject:"-Zeile!
- Denken Sie an die Leserschaft!
- Vorsicht mit Humor, Ironie und Sarkasmus!
- Kürzen Sie zitierten Text auf das notwendige Minimum!
- Benutzen Sie E-Mail!
- Geben Sie eine Sammlung Ihrer Erkenntnisse an das Netz weiter!
- Achten Sie auf die gesetzlichen Regelungen!
- Benutzen Sie Ihren wirklichen Namen, kein Pseudonym!
- Vorsicht mit Kommerziellem!
- Vorsicht mit Binaries und Multipart-Artikeln!
- "Du" oder "Sie"?

Newsgroups: Regeln

Zum Schluss:

- von Elchen, Plonken, TOFU und Trollen:
<http://www.uwe-stoeckert.de/usenet/spezial.htm>