

5.3 Suchdienste

Die bekanntesten Vertreter sind Altavista (www.altavista.digital.com), Lycos (www.lycos.com), Excite (www.excite.com), Infoseek (www.infoseek.com), Hotbot (www.hotbot.com), Fireball (www.fireball.de) und Blue Window Search (www.sear.ch/d).

Suchdienste basieren in der Regel auf speziellen Programmen, die das Internet ständig durchkämmen. Diese «Roboter», «Spider» oder «Crawler» lesen in regelmässigen Abständen alle ihnen bekannten Web-Seiten (und andere Internet-Gefässe) und registrieren sie in ihrer Datenbank. Der Benutzer greift später lediglich auf jene Datenbank zu, die die Adressen der Seiten, eine kurze Beschreibung und die darin vorkommenden Suchbegriffe enthält.

Tip Eine Reihe von Suchdiensten beschränken sich auf bestimmte Sprachregionen bzw. Top-Level-Domains (z.B. nur «.ch»-Adressen). Diese bieten den Vorteil, einen kleineren Bereich als die grossen internationalen Dienste abdecken zu müssen und daher ihren Datenbestand häufiger aktualisieren zu können. Wer deutschsprachige Inhalte sucht, sollte daher zunächst einen deutschsprachigen Suchdienst benutzen (z.B. www.fireball.de oder www.sear.ch/d). Einige internationale Suchdienste (z.B. Altavista) bieten zwar die Möglichkeit, in Suchläufe nur deutschsprachige Web-Sites einzubeziehen. An der geringeren Aktualität ändert dies jedoch nichts.

Im Wörterindex eines Suchdienstes sind in der Regel fast nie wirklich alle Wörter einer Seite enthalten. Einige Suchdienste lassen allzu häufig vorkommende Begriffe von vornherein weg («stop words»). So ergibt eine Suche nach Wörtern «you» oder «more» mit der US-Version von Altavista null Treffer, während er «du» oder «mehr» akzeptiert; deutsche Begriffe sind für einen amerikanischen Suchdienst eben etwas selteneres. Andere Suchdienste (z.B. Lycos) merken sich nur Wörter am Anfang einer Seite oder speichern nur die «wichtigsten» (= seltensten) Wörter.

An neue Internet-Adressen gelangen die Suchdienste einerseits über Eingaben von Benutzern, andererseits bahnen sich die Suchroboter ihren eigenen Weg: Enthält eine Homepage einen Link auf eine andere Web-Seite, wird der Suchroboter diesen Link ausprobieren und gelangt so auf eine ihm vielleicht noch unbekannt Seite.

Tip Jeder Suchdienst stellt seine Datenbank auf unterschiedliche Weise zusammen. Derselbe Suchbegriff wird daher in jedem Suchdienst zu anderen Ergebnissen führen. Es gibt internationale Suchdienste, die vor allem in den USA stark sind, während andere sich besonders gut zur Suche nach Informationen in einem bestimmten Fachbereich eignen. Meist lohnt es sich deshalb, mehrere Suchdienste zugleich zu benutzen.

Kein Suchdienst hat in seiner Datenbank sämtliche Seiten bzw. Dokumente des Internets verzeichnet. Einige wollen nach eigenen Angaben eine Abdeckung von 80 bis 90 Prozent erreicht haben, doch erscheint dies nicht besonders glaubwürdig. Gegen einen derart hohen Grad der Erfassung spricht schon, dass manche Seiten im Internet von Suchdiensten grundsätzlich nicht erfasst werden bzw. erfasst werden können.

So ist es Anbietern möglich, ihre Web-Seiten derart zu kennzeichnen, dass die meisten Suchroboter sie bewusst nicht festhalten (vgl. info.webcrawler.com/mak/projects/robots/norobots.html).

Manche Suchroboter erfassen Angebote zudem nur bis zu einer gewissen Ebene; Seiten, welche zu tief in einem Angebot versteckt sind, werden vom Suchroboter nicht mehr beachtet.

Auch Angebote, die den Benutzer auf eine Datenbank zugreifen lassen, können von Suchrobotern nicht richtig erfasst werden, da diese nur bis zur Suchmaske vordringen können. Die Datenbankinhalte werden sie nicht zu Gesicht bekommen, da diese in der Regel immer nur auszugsweise und erst nach Eingabe eines entsprechenden Suchbegriffs angezeigt werden. So gehen den Suchrobotern viele wichtige Inhalte verloren, zumal eine Reihe von Internet-Anbietern über ihre Web-Seiten Informationen direkt aus Datenbanken anbieten statt sie in Form fertiger Web-Seiten auf ihrem Web-Server zu lagern.

Tip Es gibt im Internet inzwischen spezielle Suchdienste für Datenbanken. Diese sind allerdings oft kostenpflichtig oder aber auf bestimmte Themenbereiche beschränkt. Relativ preiswert sind z.B. www.northernlight.com, www.intellisearchnow.com.

Für Suchdienste besteht das Internet nicht aus kompletten Angeboten, sondern nur aus vielen einzelnen Seiten. Finden die Suchdienste eine Seite, die dem Suchauftrag entspricht, spielt der Kontext des Gesamtangebots, aus dem sie stammt, keine Rolle und wird nicht berücksichtigt. Dementsprechend ist ein Suchdienst überall dort fehl am Platz, wo sich die gesuchte Information nicht auf einer einzigen Seite befindet, sondern über mehrere Seiten verteilt ist. Suchdienste prüfen alle Seiten «isoliert», nicht im Verbund. Wird nach zwei zwingenden Begriffen gesucht, von denen der eine auf der ersten, der andere auf der zweiten Seite eines Angebots zu finden ist, so wird eine Suchmaschine keine der beiden Seiten als Treffer erkennen. Immerhin kann eine von einem Suchdienst gefundene Seite einen guten Einstiegspunkt in ein Angebot zum gesuchten Thema bieten, so dass der Benutzer von dieser Seite aus das restliche Angebot erkunden kann (vgl. S. 121).

Ein Suchdienst eignet sich in erster Linie zur Klärung einer bestimmten Frage oder eines einzigen, spezifischen Sachverhalts und nicht zur Suche nach Informationen zu einem bestimmten Themenkomplex. Der Benutzer muss mehr oder weniger genau wissen, was er will. Während er in Katalogdiensten meist nur nach Themenbezeichnungen suchen kann, erlaubt ein Suchdienst die Suche nach Wörtern. Je ausgefallener und seltener diese sind, desto grösser sind die Erfolgchancen. Besonders hoch sind sie etwa bei der Suche nach ganzen Phrasen (z.B. Zitaten), Fachbegriffen (z.B. «Herzklappe») oder festen Bezeichnungen und Namen (z.B. «Maastricht», «Lufthansa»).

Tip Während Katalogdienste bei der Suche nach dem Namen «Lufthansa» im Grossen und Ganzen nur die von der Lufthansa und ihren Partnern selbst unterhaltenen Web-Angebote finden, zeigen Suchdienste auch Seiten an, die aus anderen Angeboten stammen und den Begriff «Lufthansa» enthalten. Das könnte zum Beispiel der Artikel einer Zeitung im Internet sein, der über Lufthansa berichtet. Wer «inoffizielle» Informationen über eine Firma oder zu einer Bezeichnung sucht, fährt demnach mit Suchdiensten möglicherweise besser.

Suchdienste lassen sich durch Eingabe bestimmter Suchbegriffe und Suchbefehle abfragen. Ob eine bestimmte Information gefunden wird, hängt somit entscheidend davon ab, wie gut der Suchauftrag formuliert wurde (vgl. S. 166).

Die meisten Suchdienste bieten zwei Suchmasken an. Die «einfache Suche» (auf englisch oft «simple search» genannt) ist die voreingestellte Methode: Sie erlaubt die Verwendung einfacher Suchbegriffe und einige wenige Suchbefehle. Den meisten Anwendern genügt diese Suchmaske. Demgegenüber bietet die «erweiterte Suche» oder «Expertensuche» (auf englisch oft «expert search» oder «extended search» genannt) die Möglichkeit, Suchbegriffe auf besondere Art zu verknüpfen, zusätzliche Suchbefehle einzugeben und nach bestimmten Elementen auf einer Web-Seite zu suchen. Zwischen den beiden Modi lässt sich normalerweise durch Anklicken eines entsprechenden Umschaltfeldes hin- und herschalten.

Tip Fast jede Suchmaschine verfügt über mehrere "Hilfe"-Seiten, die genau beschreiben, wie Suchaufträge formuliert und welche Suchbefehle wo und wozu verwendet werden können. Es lohnt sich, sich diese Seiten anzuschauen (vgl. S. 166).

Immer häufiger bieten Suchmaschinen nach Anzeige der ersten Treffer die Möglichkeit, den Suchauftrag weiter zu verfeinern, sei es, um die Zahl der Treffer auszuweiten, sei es, um den Suchauftrag noch etwas enger zu definieren. Eine Reihe von Suchmaschinen zeigt dazu in einem Feld den bisherigen Suchauftrag an, der dann ergänzt werden kann (z.B. Altavista, Lycos, Hotbot). Der Suchauftrag wird dann in seiner modifizierten Form von neuem gestartet.

Suchmaschinen bieten zum Teil auch die Möglichkeit, einen zweiten Suchlauf zu starten, in den jedoch nur die bereits gefundenen Treffer einbezogen werden (z.B. Lycos, Infoseek): Wurde zunächst nach dem Stichwort «New York» gesucht und dieser Suchlauf nun mit dem Begriff «musical» fortgesetzt, wird die Suchmaschine in diesem Falle nur noch die Web-Sites nach dem Begriff «musical» absuchen, in denen im ersten Durchlauf das Stichwort «New York» schon gefunden wurde.

Einige Suchmaschinen gehen noch weiter (z.B. Altavista, Excite). Sie analysieren die von ihnen gefundenen Dokumente im Internet auf Häufungen bestimmter Begriffe und schlagen dann eine Reihe von weiteren Wörtern vor, die häufig im Zusammenhang mit dem vom Benutzer bereits gewählten Suchbegriff vorkommen. Dieser kann dann auswählen, welche er in seinen Suchauftrag einbeziehen und welche er explizit ausschliessen will. Die Erfahrung zeigt allerdings, dass diese Suchhilfen in der Praxis oft

nicht sehr nützlich sind; die technische Entwicklung befindet sich hier noch ganz am Anfang.

Wer zum Beispiel in Altavista nach dem Begriff «Gates» sucht, wird eine Trefferliste mit mindestens 40'000 Seiten erhalten. Wird nun der «Refine»-Knopf (Verfeinern) angeklickt, zeigt Altavista rund 20 Begriffsgruppen mit Wörtern an, die irgendwie im Zusammenhang mit Gates stehen. Auf der ersten Zeile erscheinen Begriffe wie «microsoft», «bill» und «windows». Auf einer anderen Zeile dagegen sind Begriffe wie «programmable», «cmos» oder «fpga» zu finden und auf einer weiteren «teledesic», «mccaw» und «wireless» (diese Begriffe ändern sich übrigens im Laufe der Zeit; manche kommen hinzu, manche verschwinden).

Wer nicht nach Bill Gates und Microsoft sucht, sondern nach sogenannten «Gatter»-Bauteilen (englisch «gates»), wie sie in der Elektronik verwendet werden, kann nun die Begriffe der ersten Zeile («microsoft» etc.) ausschliessen («exclude» wählen) lassen, während er die Zeile mit den Elektroniker-Begriffen («cmos» etc.) ausdrücklich in seinen Suchauftrag einschliesst («require» wählen) und einen neuen Suchlauf startet. Die Zahl der Treffer wird auf weniger als einen Zehntel reduziert.

Auch wer nach Bill Gates sucht, kann seine Suche mit Hilfe der Begriffsgruppen in eine bestimmte Richtung lenken. Die Begriffsgruppe mit «teledesic» wird beispielsweise vorwiegend Internet-Adressen anzeigen, unter denen etwas über Bill Gates Investment in die Satelliten-Kommunikationsfirma Teledesic zu lesen ist.

Andere Suchmaschinen (z.B. Excite) bieten mit Hilfe ähnlicher Verfahren die Möglichkeit, zu jedem Treffer weitere Adressen anzuzeigen, die ähnliche Dokumente enthalten. Hat der Benutzer ein Dokument gefunden, das seinen Vorstellungen entspricht, sucht ihm die Suchmaschine weitere Dokumente heraus, die er wahrscheinlich auch verwenden kann, weil in ihnen die selben Wortkonstellationen vorkommen.

Einige Suchmaschinen (z.B. Lycos, Infoseek) zeigen neben den Treffern auch eine kurze Liste mit Themen an, von denen das eine oder andere dem entspricht, was der Benutzer sucht. Wer zum Beispiel nach «gates» gesucht hat, dem zeigt Lycos (links oben) «bill gates», «gates iii», «crusher gates» und «beverly crusher gates» an. Wer auf die ersten beiden Begriffe klickt, lässt Lycos explizit nach Seiten über den Microsoft-Chef suchen. Wer dagegen die beiden anderen Begriffe anwählt, erhält Adressen zur Schauspielerin Gates McFadden und ihrer bekanntesten Rolle, die der Schiffsärztin Beverly Crusher aus der TV-Serie «Raumschiff Enterprise».

LESEPROBE AUS:
Rosenthal, David:

Infopool Internet : Methoden, Tricks und Quellen der Profis zur effizienten Recherche /
David Rosenthal. - Zürich : Orell Füssli, 1998
ISBN 3-280-02458-7 <http://www.insider.ch/ipd/recherche>
Die Vervielfältigung und Publikation ist nicht erlaubt.

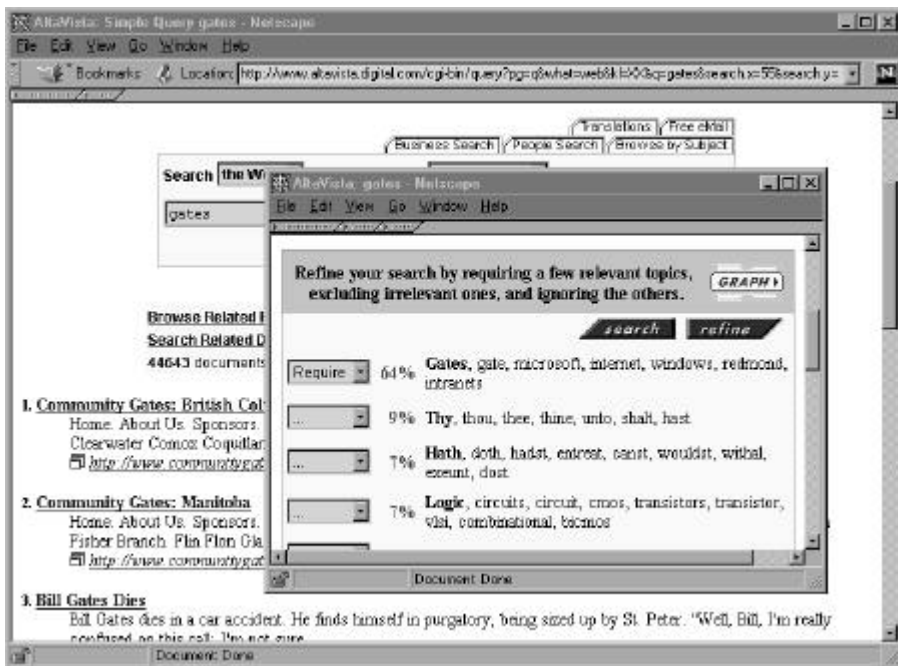


Abbildung 14: Die meisten Suchaufträge müssen ein oder mehrmals verbessert werden (hier mit der Refine-Funktion von Altavista).

Einige Suchdienste verweisen nach jedem Suchlauf auf Suchmöglichkeiten in anderen Verzeichnissen und Datenbanken. So kann mit einem Mausklick meist nach Büchern zum Suchbegriff gesucht werden, nach Bildern, Web-Site-Rezensionen, nach Adressen, Nachrichten oder privaten Homepages (z.B. Lycos, Altavista, Excite, Infoseek). Einige bieten auch Sucheinschränkungen in Form eines Links zum Anklicken an (z.B. Lycos); der bestehende Suchlauf wird damit erneut durchgeführt, jedoch nur in den Titelzeilen. Einige Suchmaschinen können zu jedem Treffer auch alle weiteren, registrierten Seiten unter derselben Adresse anzeigen (z.B. Lycos «Find related Web pages»).

Tip Die meisten Suchdienste bauen ihr Angebot angesichts des recht grossen Konkurrenzdrucks laufend aus. Es kann sich daher lohnen, die gängigen Suchdienste von Zeit zu Zeit zu besuchen und nach neuen Funktionen Ausschau zu halten.

Der Erfolg eines Suchlaufs im Internet hängt neben der Formulierung des Suchauftrags auch davon ab, wie die gefundenen Treffer angezeigt werden. Benutzer von Suchmaschinen schauen sich in der Regel immer nur die ersten paar angezeigten Adressen an.

Die wenigsten Benutzer sind sich jedoch bewusst, dass die verschiedenen Suchdienste ihre Treffer auf ganz unterschiedliche Art und Weise sortieren und anzeigen («Ranking»). Die einen Systeme zeigen zum Beispiel zunächst all jene Adressen an, bei denen der Suchbegriff bereits im Titel der Seite vorkommt. Andere führen zuerst jene Seiten auf, in denen die meisten der gesuchten Begriffe vorkommen. Die Feinsortierung wird dann oft nach dem Anteil der Suchbegriffe am Text des Gesamtdokumentes vorgenommen. Ein kurzer Text, der alle Suchbegriffe enthält, schneidet somit besser ab als ein langer Text, in dem ebenfalls sämtliche Suchbegriffe vorkommen.

Einige Suchmaschinen gewichten zudem bestimmte Seitenteile wie Titel, den Anfang einer Seite und Überschriften stärker. Auch die Suchbegriffe können einen Einfluss haben: Wird von mehreren Suchbegriffen ein selteneres Wort gefunden, so wirkt sich das mitunter stärker auf ein gutes Ranking aus als die Fundstelle eines im Internet gebräuchlicheren Begriffs.

Altavista arbeitet beispielsweise mit einem etwas komplizierten Punktesystem: Die Seite mit den meisten Punkten wird als erste aufgeführt. Dabei bringt ein seltener Suchbegriff mehr Punkte ein als ein Begriff, der in den verzeichneten Dokumenten relativ häufig vorkommt. Zudem ist die Fundstelle des Begriffs von Bedeutung: Wird der Begriff innerhalb der ersten acht Wörter des Titels gefunden, so verfünffacht sich seine Punktezahl (Annahme: Suche mit einem Suchbegriff). Wird er im restlichen Teil des Titels gefunden, so wird die Punktezahl mit vier multipliziert. Auf dritter Position folgen jene Dokumente, in denen der Suchbegriff mindestens zwei Mal im Text vorkommt (Multiplikation mit 1.5), und schliesslich jene, in denen er nur noch einmal auftaucht (keine Multiplikation). Kommt der Suchbegriff in Titel und Text vor, hat das keinen Einfluss auf die Sortierung.

Einige Suchdienste erlauben es dem Benutzer, selber Sortierkriterien zu wählen (so etwas bei Altavista durch Eingabe eines Selektionsbegriffs im erweiterten Suchmodus: Als erstes werden nur Dokumente mit dem Selektionsbegriff angezeigt). Andere erlauben es, mehrere gefundene Adressen derselben Web-Site in einem Treffer zusammenzufassen; will der Benutzer alle Adressen sehen, kann er sich diese jeweils anzeigen lassen (z.B. Exci-