Suchen und Finden im Internet

Der Umfang und die Vielfalt der im Internet verfügbaren Informationen hat mit der Entstehung des World Wide Web explosionsartig zugenommen. Wer nicht nur mal „just for fun“ durchs Netz surft, sondern die Ressourcen des Internet für seine Arbeit, politische Aktivitäten, das eigene Hobby, zur Kontaktaufnahme oder andere Dinge nutzen möchte, hat meistens relativ wenig Zeit zur Verfügung, in dem riesigen Informationsangebot das Gesuchte zu finden. Hunderttausende Internet-Nutzer (Organisationen, Firmen, Privatpersonen) veröffentlichen Informationen im Netz, ohne daß sie das irgendwo anmelden und dabei mehr als technische Festlegungen der benutzten Internet-Protokolle und gesetzliche Bestimmungen beachten müßten. Es gibt Werkzeuge und Methoden, diese Informationsflut zu bewältigen, statt sich resigniert treiben zu lassen. Einige davon werden im folgenden Artikel vorgestellt.

Überblick

Zunächst eine kurze Beschreibung der für die Suche nach Informationen zur Verfügung stehenden „Werkzeuge“. Die Indizes („Suchmaschinen“) und die Kataloge sind darunter die bekanntesten und werden am häufigsten genutzt. Sie erreichen so hohe Zugriffszahlen pro Tag, daß der Aufbau solcher Dienste kommerzielle Gewinne verspricht. Allerdings ist dafür auch ein hoher Aufwand an Hard- und Software (bei den Indizes) bzw. redaktioneller Arbeit (bei den Katalogen) erforderlich. Sie werden deshalb von kommerziellen Unternehmen betrieben, einige wenige auch von Einrichtungen aus dem akademischen Umfeld.

Indizes (Suchmaschinen)

„Robots“ oder „Spider“ genannte Programme durchsuchen automatisch rund um die Uhr die wichtigsten Internetdienste (WWW, News, Gopher u.a.) nach neu veröffentlichten oder geänderten Dokumenten. Im WWW ist das möglich, indem von bekannten Dokumenten ausgehend alle darin enthaltenen Hyperlinks zu weiteren Dokumenten verfolgt werden, dann wieder die dort enthaltenen Verweise zu den nächsten usw. Die jeweils gefundenen Objekte werden eingesammelt („heruntergeladen“), vollständig oder teilweise analysiert und das Ergebnis in Datenbanken gespeichert. Im Idealfall würde dabei ein Volltextindex aller Internetressourcen entstehen.

Die Benutzer können in diesen Datenbanken durch Suchanfragen recherchieren und erhalten als Antwort eine Liste von Verweisen auf alle in der Datenbank enthaltenen Dokumente, die ihren Suchkriterien entsprechen.

eine (unvollständige) Liste:

Indizes:

AltaVista	http://altavista.digital.com/
Excite	http://www.excite.com/
HotBot	http://www.hotbot.com/
Infoseek	http://www.infoseek.com/
	http://ultra.infoseek.com/
Lycos	http://www.lycos.com/
Opentext	http://index.opentext.net/
Webcrawler	http://webcrawler.com/
Yahoo!	http://www.yahoo.com/

speziell für Newsgruppen:

Dejanews	http://www.dejanews.com/

deutsche Indizes:

Aladin	http://www.aladin.de/
Eule	http://www.eule.de/
Flipper	http://flp.cs.tu-berlin.de/flipper/
Excite	http://de.excite.com/
Focus Netguide	http://www.netguide.de/
Kolibri	http://www.kolibri.de/
Lycos	http://www.lycos.de/
Nathan	http://www.nathan.de/

Um beurteilen zu können, wozu Indizes taugen und wozu nicht, ist etwas Hintergrundwissen über die Funktion dieser Dienste erforderlich.

Hinter dem Begriff Suchmaschine verbirgt sich in der Regel ein großer Hardwarepark, d. h. große Cluster von (miteinander vernetzten) leistungsfähigen Rechnern. Jeweils eine Gruppe davon ist für eine der anfallenden Aufgaben von der Informationsbeschaffung bis zur Beantwortung von Suchanfragen zuständig (Sammeln neuer Dokumente, periodische Überprüfung aller erfaßten Dokumente auf Änderungen, Herunterladen von Dokumenten, Analyse und Indizierung, Speicherung in der Datenbank, Bearbeitung der Suchanfragen).

Für die Speicherung in der Datenbank müssen die gesammelten Texte entsprechend aufbereitet, d.h. analysiert und indiziert werden. Die dazu verwendeten Verfahren unterscheiden sich von Index zu Index und bleiben Firmengeheimnis. Einige speichern und analysieren den gesamten Text, andere nur Überschriften, Verweise, die ersten Zeilen. Es werden nicht nur alle enthaltenen Wörter registriert, sondern auch deren Position (im Titel, in Überschriften, im Text, in Hyperlinks, ihre Nachbarschaft zu anderen Wörtern, Abstand zu anderen Suchbegriffen ...), um damit ihre Bedeutung im Dokument ermitteln zu können. Zusätzlich speichern manche die Dateinamen der in Webseiten eingebundenen multimedialen Objekte (Bilder, Java-Applets etc.). Mathematisch-statistische Verfahren erlauben es, zu einem gegebenen Dokument ähnliche Dokumente in der Datenbank zu finden und dem Nutzer vorzulegen (*Excite, Infoseek*).

Je detaillierter und aufwendiger die Analyse ist, desto genauer können bestimmte Kriterien später bei der Suche erfragt werden.

Die größten Suchmaschinen halten in ihren Datenbanken Informationen über Millionen von Webseiten (*Alta Vista, Hotbot, Lycos, Excite* – jeweils etwa 60 Mill.). Je nach Genauigkeit der Anfrage erhält man eine unterschiedlich große Anzahl von Dokumenten, die die gesuchten Informationen enthalten, neben vielen anderen, die zwar formal die Suchanfrage erfüllen, aber nicht zum Thema passen. Deshalb ist es von Vorteil, wenn Suchmaschinen dem Nutzer die Ergebnisse sortiert nach ihrer Relevanz präsentieren (die nach vorgegebenen Kriterien berechnet wurden). Das bewahrt ihn vor dem aufwendigen Durchsuchen unzähliger Webseiten. Außerdem liefert das auch Nutzern, die mit den Abfragemöglichkeiten der jeweiligen Suchmaschine noch nicht so gut vertraut sind, brauchbare Ergebnisse schon bei einfachen (unscharfen) Suchanfragen. Für die Berechnung der Relevanz verwenden die Firmen ebenfalls unterschiedliche Verfahren (ausführlichere Informationen dazu z.B. in [1] und [4]).

Folgendes finden Suchmaschinen nicht:
- Dokumente auf Servern, auf die noch keine Verweise zeigen;
- Dokumente, die erst nach der letzten Aktualisierung der Datenbank veröffentlicht oder geändert wurden (die Zeiten, die die einzelnen Dienste benötigen, um ihren gesamten Datenbestand einmal zu aktualisieren, sind unterschiedlich);
- durch Paßworte oder Firewalls geschützte Dokumente;
- Dokumente in Bereichen, die vom Server-Betreiber für Suchroboter gesperrt wurden oder die eine Registrierung verlangen;
- von CGI-Skripten oder Datenbank-Gateways dynamisch erzeugte Dokumente (d.h. alle in Datenbanken gespeicherten Informationen - außer wenn diese offline als Webseiten aufbereitet werden - können nicht gefunden werden)

Einen „besten“ Index gibt es nicht. Die Betreiber verbessern ständig die eingesetzte Hard- und Software, um die einzelnen Teilaufgaben (siehe oben, z.B. die Analyse der Dokumente) schneller und intelligenter ausführen zu können, denn sie stehen in Konkurrenz zueinander. Es ist ratsam, diese Entwicklung zu verfolgen und sich in Abständen immer wieder neu für ein oder zwei persönlich bevorzugte Anbieter zu entscheiden.

Unterschiede bei den Indizes:
- Anspruch auf Vollständigkeit oder Spezialisierung auf Dokumente in einer bestimmten Sprache, zu einzelnen Teilgebieten, aus bestimmten Internetdiensten;
- in den Verfahren bei der Suche nach neuen oder geänderten Informationen;
- in der Aktualität der gespeicherten Dokumente (in der Dauer der Aktualisierungszyklen);
- beim Aufwand für die Analyse der Dokumente;
- im Umfang und der Leistungsfähigkeit der Datenbanken (Antwortzeiten);
- in der Flexibilität der Abfragesprache, in der die Benutzer die Suchanfragen formulieren können;
- in der Unterstützung der Nutzer bei der Auswertung der Suchergebnisse (Relevanzberechnung, weitere Abfragen zur Eingrenzung der Treffer usw.)

Kataloge

Im Unterschied zu den Indizes werden hier die im Internet aufgespürten Dokumente nach Themengebieten geordnet und als hierarchisch geordnete Verzeichnisse zur Verfügung gestellt. Auch diese Dienste wer-

den überwiegend von Firmen aufgebaut. Die Benutzung ist kostenlos, Gewinne sollen durch den Verkauf der Werbeflächen auf den Webseiten erwirtschaftet werden. Den Hauptanteil der Suche und der thematischen Einordnung leisten Redaktionen, einige ergänzen das durch automatische Suche oder Sortierung mit Robotern und die Möglichkeit, daß die Anbieter von WWW-Seiten ihre Angebote selbst eintragen können.

eine (unvollständige) Liste:

Kataloge:

Tradewave Galaxy	http://galaxy.einet.net/
WWW Virtual Library	http://www.w3.org/vl/
Yahoo!	http://www.yahoo.com/

deutsche Kataloge:

„deutsche Datenquellen“ (deutscher Ast der WWW Virtual Library)
http://www.rz.uni-karlsruhe.de/Outerspace/VirtualLibrary/

DINO	http://www.dino-online.de/
Web.DE	http://web.de/
Yahoo!	http://www.yahoo.de/

Die Anzahl der in den Katalogen erfaßten Dokumente ist wesentlich geringer als bei den Indizes (Schätzung bei Yahoo! : 250000 ... 500000). Dafür ist auch wegen der Vorsortierung und Kommentierung der Inhalte durch Redaktionen eine größere Trefferquote bei der Suche zu erwarten. Bei den meisten Katalogen wird ebenfalls eine Indizierung der Dokumente vorgenommen. Es werden Eingabemasken für die Formulierung von Suchanfragen im Datenbestand des gesamten Verzeichnisses angeboten.

Die großen Anbieter (siehe Liste) haben den Anspruch, das gesamte Internetangebot thematisch zu ordnen. Diese Verzeichnisse eignen sich deshalb gut als Einstiegspunkt einer Suche. Daneben sollte man die spezialisierten Kataloge nicht vergessen.

Meta-Suchhilfen

Wegen der Unterschiede in Anspruch und Qualität der Indizes und Kataloge ist es oft ratsam, bei einer Suche nach bestimmten Informationen, mehrere dieser Dienste zu nutzen. Meta-Suchmaschinen sollen die oft zeitraubende Arbeit der Einarbeitung in die unterschiedliche Abfragesyntax der Suchmaschinen und die Auswertung der Suchergebnisse erleichtern. Sie halten keine eigenen Datenbestände.

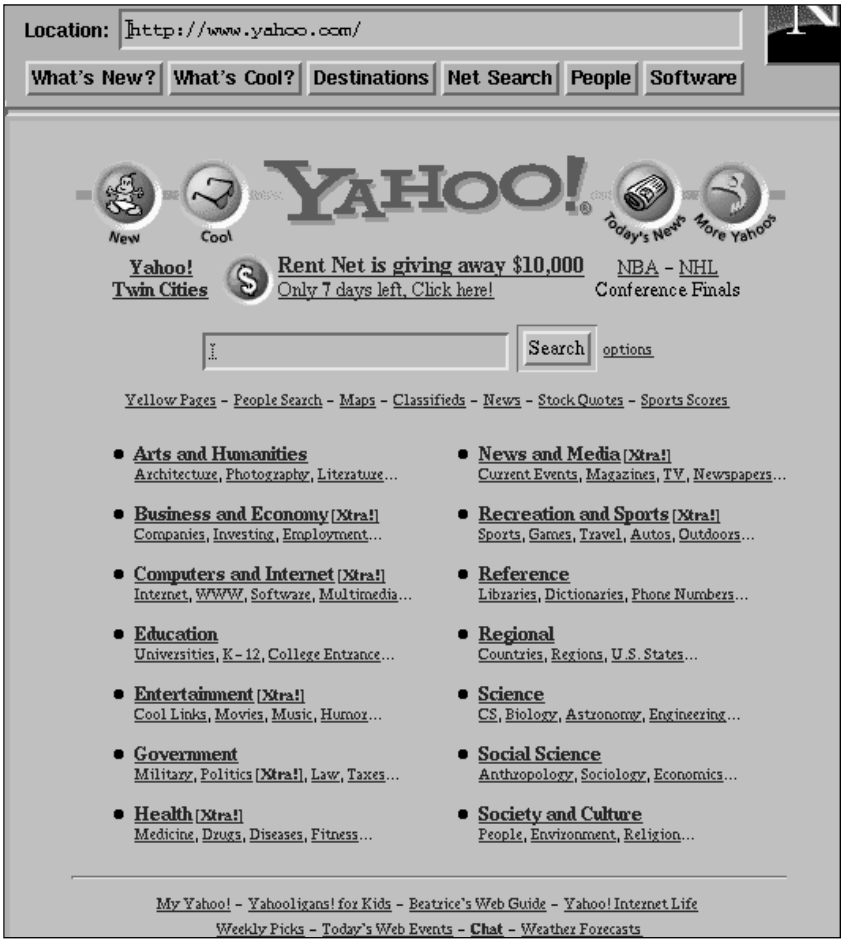
Es gibt inzwischen fast eine inflationäre Tendenz beim Angebot solcher Werkzeuge (auch Meta-Meta-Suchmaschinen).

Meta-Suchhilfen:

CUSI	http://web.nexor.co.uk/public/cusi/cusi.html
CUSI (Deutschland)	http://www.unix-ag.uni-siegen.de/search/
Internet Sleuth	http://www.isleuth.com/
Search.com	http://www.search.com/
Metacrawler	http://www.metacrawler.com/
MetaGer	http://meta.rrzn.uni-hannover.de/

Besprechungsdienste

Der Datenbestand dieser Dienste ist im Vergleich zu dem der Kataloge noch geringer, denn hier geht es nicht um Vollständigkeit, sondern um Qualität: Redaktionen wählen innerhalb der verschiedenen Sachgebiete die ihrer Meinung nach besten Internet-Angebote aus. Die Kriterien dieser Auswahl werden mitveröffentlicht. Wenn man zu einem bestimmten Thema schnell einen



Überblick über wichtige Angebote finden will, sind solche Dienste dafür gut geeignet.

Besprechungsdienste:

Argus Clearinghouse of Subject Oriented Guides to Internet Resources	http://www.clearinghouse.net/
Lycos Top 5%	http://www.pointcom.com/categories/
Magellan	http://magellan.mckinley.com/
Webtip	http://www.webtip.de/

Weitere Suchhilfen

Kataloge und Indizes (und darauf aufbauende Dienste) sind zwar die wichtigsten Suchwerkzeuge, sie eignen sich aber nicht für alle Recherchen. Einerseits erfassen Suchmaschinen nicht alle Ressourcen (siehe oben), andererseits gibt es Internetdienste, die genau für einen speziellen Zweck geschaffen wurden und deshalb geeigneter sind. Für fast alle diese Dienste sind WWW-Gateways geschaffen worden, d.h. sie sind mit einem WWW-Browser nutzbar. Das ist bis auf wenige Ausnahmen günstiger als die Benutzung der speziell dafür geschaffenen Programme (Clients). Für Dienste, wie FTP und Gopher, die vor dem WWW entstanden, existieren eigene Suchhilfen (FTP: *Archie*, Gopher: *Veronica*). Im Abschnitt *Suchthemen: einige Beispiele* tauchen solche Dienste noch einmal auf und werden hier deshalb nicht weiter beschrieben.

Einfache Suche

In diesem Abschnitt geht es um Recherchen, die ohne größeren Aufwand (Nachdenken über eine sinnvolle Recherchestrategie, Einarbeitung in die Abfragesyntax einer der Suchmaschinen usw.) durchgeführt werden können. Mit etwas Wissen über die richtige Methode kommt man schnell zum Ziel, kann an das Ergebnis jedoch keine großen Anforderungen stellen („quick and dirty"-Recherche).

Zur Auswahl stehen hier vor allem die beiden wichtigsten Suchhilfen: Indizes und Kataloge.

Suche in einem Index

Als Beispiel soll hier die Suche nach dem Wort „Transrapid" dienen (Suche nach Informationen zum umstrittenen Projekt Magnetschwebewahn). Die Suche bei *Excite* (weltweit) ergibt 574, bei *Hotbot* 909 und bei *AltaVista* 400 Fundstellen - begleitet von Werbung für Pampers, Rasenmäher usw. Diese Zahlen allein sagen nichts aus über die Verwertbarkeit der gefundenen Dokumente im Sinne unserer Suche. Es wird jedoch klar, daß die Auswertung aller Fundstellen viel Zeit in Anspruch nehmen würde bzw. eher undurchführbar ist. Da die besten Suchmaschinen eine Relevanzbewertung vornehmen (*Excite* gibt z.B. zu jeder Fundstelle die berechnete Relevanz mit an) und die gefundenen Dokumente geordnet nach den berechneten Werten anzeigen, sind bereits unter den ersten 10 Dokumenten genügend Treffer, die für einen Einstieg in das Thema reichen würden (z.B. die offizielle Transrapid Homepage). Für eine Recherche nach den Argumenten der in der Auseinandersetzung um das Projekt beteiligten Interessengruppen reicht das wahrscheinlich nicht aus. Dazu müßte man erweiterte Suchmöglichkeiten anwenden, um möglichst effektiv zu Ergebnissen zu kommen.

Hinweise zur einfachen Stichwortsuche:

- Einen oder mehrere möglichst treffende Suchbegriffe verwenden, um die Anzahl der nicht verwertbaren Suchergebnisse gering zu halten. Keine allgemeinen Begriffe, wie Kunst, Pflanzen usw,. angeben.
- Zur Klein- oder Großschreibung: kleingeschriebene Suchbegriffe liefern sowohl klein- als auch großgeschriebene Wörter, großgeschriebene Begriffe nur große Wörter. Ausnahme: bei Eigennamen ist die Originalschreibweise angebracht.
- Wer nur Dokumente in einer bestimmten Sprache sucht, sollte versuchen, mit einem auf diese Sprache spezialisierten Index oder – falls vorhanden – dem entsprechenden „Ableger" der großen Indizes zu beginnen (z.B. für deutsche Dokumente mit der deutschen Ausgabe von Excite). Für eine Suche in den großen Indizes, die natürlich überwiegend englischsprachige Dokumente indiziert haben, muß man aussagekräftige englische Suchbegriffe formulieren, damit die Suche bei spezielleren Themen nicht ins Leere läuft (keine Dokumente gefunden, weil es keine deutsprachigen zum Thema gibt).

Suche in einem Katalog

Die einfache Suche in einem thematischen Verzeichnis, wie *Yahoo!* oder *Web.DE,* ist nichts weiter als eine Suche, die ausgehend von einer geeigneten Hauptkategorie auf der Homepage des Katalogs durch fortschreitende Eingrenzung des Themas für den Suchbegriff irgendwann auf eine Seite führen muß, die Links auf Dokumente zum gefragten Thema enthält.

Am Beispiel der Suche zu Informationen zum „Transrapid" müßte zunächst die geeignete Hauptkategorie gesucht werden, bei *Yahoo!* z.B. „Business and Economy", bei *Web.DE* z.B. Wirtschaft. Die weiteren Unterverzeichnisse lauten dann bei *Yahoo!*: „Transportation/Trains and Railroads/High Speed Rail", bei *Web.DE*: „Verkehr/Bahn".

Ein Problem der Kataloge ist (ähnlich wie bei Bibliotheken): ein bestimmtes Thema kann unter mehreren Oberbegriffen einsortiert werden. Die Ansichten darüber sind verschieden. Welchen Begriff haben jedoch die Redakteure des jeweiligen Katalogs gewählt ? Gute Kataloge haben Querverweise zu verwandten Themen. Wenn man sich unsicher ist, wo ein spezielles Thema einsortiert sein könnte, hilft am

besten eine einfache Stichwortsuche auf der Hauptseite des betreffenden Katalogs. Die Stichwortsuche nach „transrapid“ ergibt bei *Yahoo!*, daß man auch bei einer Suche unter „Regional/Countries/German/Transportation“ fündig werden könnte. Bei *Web.DE* wäre auch eine Suche unter „Organisationen“ (z.B. Aussagen von Parteien und Organisationen zum Thema), „Private Homepages“ (Meinungen engagierter oder betroffener Leute) oder unter „Regional“ (Hamburg, Berlin usw.) denkbar gewesen - führt aber zu keinem Ergebnis. Das Suchbeispiel (zunächst ja vor allem ein deutsches Thema) zeigt auch, daß die Suche zumindest auf deutschen Suchmaschinen oder in einem deutschen Katalog begonnen werden sollte.

Erweiterte Suche

Fast alle Indizes und Kataloge bieten erweiterte Suchmöglichkeiten an, die eine effektivere Suche gestatten (im Vergleich zur einfachen Suche: Verringerung der Trefferanzahl, dafür aber höhere Relevanz der einzelnen Treffer). Da die Anbieter unterschiedliche Methoden für die Analyse und Aufbereitung der im Internet gefundenen Dokumente verwenden, unterscheidet sich auch der Umfang und die Flexibilität der angebotenen Suchmöglichkeiten. Aus Platzgründen können hier die Dienste mit ihren Angeboten nicht einzeln vorgestellt werden. Es soll jedoch wenigstens benannt werden, „was es alles gibt“. Genauere Informationen über die Syntax der Abfragesprache findet man jeweils unter dem Verweis „Help“/„Hilfe“ auf der Hauptseite der Suchdienste.

Die wichtigsten erweiterten Techniken werden im folgenden kurz erklärt:

Eingabe mehrerer Suchbegriffe (unscharfe Suche):
In der Eingabemaske werden einfach mehrere Suchbegriffe, die in den Dokumenten auftauchen können, hintereinander aufgeführt. Es müssen dann nicht in jedem Dokument alle Begriffe vorkommen. Je nachdem, wieviel Begriffe in den Dokumenten vorhanden sind, werden diese dann bei verschiedenen Diensten (je nach Verfahren zur Relevanzberechnung) als Treffer bewertet oder nicht bzw. in anderer Reihenfolge angezeigt.

Logische Verknüpfung von Suchbegriffen (trennscharfe Suche):
Einige Dienste unterstützen Verknüpfungsoperatoren, wie AND, OR und NOT. Durch gut überlegte Verknüpfung ist es möglich, die Treffermenge deutlich einzugrenzen. Dazu gehört auch die Möglichkeit, durch die Operatoren „+“ bzw. „-“ Wörter zu kennzeichnen, die in den Dokumenten zwingend vorkommen müssen („+“) bzw. nicht vorkommen dürfen („-“). Das ist nützlich bei Suchbegriffen mit Mehrfachbedeutung, wie z.B. Schloß (Königsschloß, Türschloß).

Suche nach Wortteilen:
Ist diese Möglichkeit nicht vorhanden, werden z.B. entweder Dokumente, die einen Suchbegriff nur in der Einzahl (z.B. *Maske*) enthalten oder die, in denen er nur in der Mehrzahl vorkommt (*Masken*), nicht gefunden. Durch die Eingabe von Jokerzeichen (*Maske**) finden sich dann alle Dokumente in der Trefferliste. Diese Möglichkeit heißt Trunkierung (engl. truncation oder stemming). Die Suche nach Wortteilen inmitten eines Wortes ist ein weiterer Fall. Bei einigen Diensten kann man wählen, ob die eingegebenen Begriffe als Substrings oder als komplette Wörter behandelt werden sollen.

Abstandsoperatoren:
Es gibt Suchdienste, die bei der Aufbereitung der Dokumente für das Abspeichern in der Datenbank auch registrieren, welche Wörter sich in der näheren Umgebung eines bestimmten Wortes (Abstand bis zu 20 ... 100 Wörter) befinden (Wortumfeld). Der Sinn besteht in der Annahme, das Dokumente, in denen Suchbegriffe nahe beieinander stehen, als Treffer zum konkreten Thema relevanter sind als jene, bei denen das nicht der Fall ist. Für die Formulierung solcher Bedingungen gibt es die Operatoren NEAR (nahe) und ADJ (adjacent - entfernt).

Phrasensuche:
Besonders bei Bezeichnungen und Namen (von Personen, Produkten, Firmen usw.), die aus mehreren Wörtern bestehen, wird die Suche effektiver, wenn das in der Suchanfrage gekennzeichnet werden kann (im Prinzip ein Spezialfall des Operators NEAR). Dazu werden die zusammengehörigen Wörter in Anführungszeichen eingeschlossen (z.B. *„Palast der Republik“*).

Unterstützung von Umlauten und Sonderzeichen:
Nicht alle Server unterstützen das. Unbedingt notwendig ist diese Eigenschaft jedoch besonders bei der Suche nach deutschen Begriffen.

Einschränkung der Treffer auf bestimmte Attribute oder Elemente der Dokumente:
Die Suche nach Informationen über eine Webseite (bei Autor, Titel, Entstehungsdatum, Art des Dokuments, z.B. private Homepage, Benutzerinfo oder Produktbeschreibung) ist schwierig, da es keine eindeutigen Festlegungen darüber gibt, wie und wo sie zu hinterlegen sind. Dienste, die eine Suche nach Attributen unterstützen, ermöglichen das entweder über eine bestimmte Syntax (bei den folgenden Beispielen die Schlüsselwörter in Klammern) oder über Eingabemasken. Bei einigen Suchdiensten kann z.B. nach folgenden Elementen gesucht werden: Titel (*titel:*), Text in Hyperlinks (*anchor:*), Java-Applets (*applet:*), Bildern (*image:*), URL (*url:*), Servername (*host:*), Domain (*domain:*), nach dem Entstehungsdatum, nach bestimmten Dateitypen. Ebenso kann bei Artikeln aus Newsgruppen

nach dem Namen der Newsgruppe, dem Namen des Autors, seiner E-Mail-Adresse und der Subject-Zeile gesucht werden.

Zusammenfassung von Suchaspekten mit Klammern: Nur dann, wenn bestimmte Teilaspekte einer Suche zusammengefaßt werden können, sind in manchen Fällen sinnvolle Suchanfragen formulierbar. Die Regeln lauten wie in mathematischen Ausdrücken. Beispiel: *software AND (mac OR unix AND NOT windows) AND graphic*.

Suchstrategien

Man kann schneller zum Ziel kommen, wenn man sich vor einer Recherche etwas Zeit für einige Vorüberlegungen nimmt und danach dann die richtigen Suchwerkzeuge und Einstiegspunkte auswählt.

Wichtige Fragen:

- Sind die gesuchten Informationen tatsächlich schneller im Internet zu finden oder doch eher per Telefon oder Fax ?
- Wer kommt als Anbieter der gesuchten Informationen in Frage? Einige Beispiele: Autor/Herausgeber von Veröffentlichungen, Hersteller/Händler von Produkten, bei politischen Dokumenten: Parteien, Organisationen, Initiativen aus dem thematischen Umfeld (z.B. mit Kommentaren, Stellungnahmen, Ankündigungen von Aktionen), unmittelbar Betroffene, bei Dokumenten aus dem wissenschaftlichen Bereich: Forschungsinstitutionen mit diesem thematischen Schwerpunkt, wissenschaftliche Gremien, Veranstalter von Konferenzen, bei Kulturereignissen: die Einrichtungen und die Orte, in denen sie stattfinden, bei (fast allen) Themen: die Medien. Die richtige Idee an dieser Stelle ist fast schon die halbe Miete.
- Was sind die geeigneten Suchbegriffe? Sie sollten zunächst so speziell und themenspezifisch wie möglich formuliert werden (besonders bei der Suche in Spezialkatalogen). Erst wenn diese Suche fehlschlägt, kann man die Begriffe verallgemeinern (vom Speziellen zum Allgemeinen). In welcher Sprache werden die meisten Informationen zum betreffenden Thema veröffentlicht sein (meist in Englisch)? Die Suchbegriffe müssen darauf abgestimmt sein.

Suchthemen: einige Beispiele

In diesem Abschnitt werden zu Themen, die bei Recherchen oft eine Rolle spielen, einige wichtige Quellen und Dienst genannt, die für die Suche herangezogen werden können.

Listen für den Einstieg in Recherchen finden Sie auch auf dem WWW-Server du HU unter:
http://www.hu-berlin.de/outside/

Firmen, Organisationen, Hochschulen ...

Oft geht es um die Frage: Ist der/die Gesuchte überhaupt im Internet mit eigenen Informationen (d.h. in der Regel mit einer eigenen Homepage) präsent ?

An dieser Stelle muß man nicht nur raten - obwohl das in viellen Fällen zum Ziel führt. Der Name des gesuchten Anbieters spiegelt sich häufig im Domainnamen wider: http://www.microsoft.de/, http://www.hu-berlin.de. Dort werden jedoch oft Abkürzungen oder anderslautende Namen (weil der erste Name z.B. schon als Domain vergeben war) verwendet, dann wird die Suche schwierig. Das NIC (Network Information Center - die Einrichtung, bei der Domainnamen beantragt werden müssen) bzw. die nationalen Ableger davon bieten auf Ihren Homepages eine Suche in ihren Registrierungsdatenbanken an.

Für Deutschland lautet die Adresse:
http://www.nic.de/Domains/Domains.html,

beim InterNIC:
http://rs.internic.net/cgi-bin/whois/.

Software

Das Problem der Softwaresuche unterscheidet sich etwas von anderen Suchthemen. Software, die von (FTP- oder WWW-) Servern im Internet heruntergeladen werden kann, ist dort in den seltensten Fällen unter ihrem Titel abgelegt, sondern unter einem kryptischen Dateinamen (der z.B. durch die Notwendigkeit erzwungen wurde, in einen Dateinamen (8 Zeichen unter MS-DOS) so viel Information wie möglich zu packen zu müssen, z.B. twsk20b.zip für Trumpet Winsock Version 2.0b). Nur selten kennt man diesen Namen. Die folgenden Hinweise sind natürlich jeweils nur eine Auswahl der vorhandenen Möglichkeiten.

(1) Wer Software für einen bestimmten Zweck sucht (z.B. ein FTP-Tool für Windows 95), sich aber noch nicht für eine konkrete Software entschieden hat, hat mehrere Möglichkeiten:

(1a) die Suche in einem Besprechungsdienst, z.B. bei Lycos Top 5% unter “Computers und Software”-”Archives” usw. Auf einigen der dort eingetragenen Servern findet man auch Beurteilungen zu Software.

(1b) den Einstieg über einen Katalog (meistens Hauptkategorie “Computer”);

(1c) die Suche auf einem WWW- bzw. FTP-Server mit nutzerfreundlicher graphischer Oberfläche (man könnte diese als Kataloge bezeichnen, die auf Software spezialisiert sind), z.B.
http://www.leo.org/archiv/
http://www.download.com/
http://www.jumbo.com/

(2) Wenn nach einem konkreten Softwarepaket gesucht wird (und der Dateiname nicht bekannt ist,

unter dem sie auf FTP-Servern zu finden ist), gibt es z.B. folgende Wege:

(2a) die Nutzung von Suchabfragen in Katalogen (am besten schon in einer speziellen Unterkategorie),

(2b) die Stichwortsuche auf Servern vom Typ (1c).
Weitere Adressen:
Virtual Software Library (Ableger in Chemnitz)
http://www1.tu-chemnitz.de/cgi-bin/vsl-front/;
ASK-SINA
http://www.ask.uni-karlsruhe.de/asksina/sina.html

(3) Für die eigentliche Suche auf anonymen FTP-Servern (nur wenn zumindest ein Teil des Dateinamens bekannt ist) gibt es den Dienst Archie. Ähnlich wie die Suchmaschinen durchsuchen die Archie-Server weltweit die Verzeichnisse der ihnen bekannten FTP-Server, speichern die gefundenen Informationen in einer Datenbank und stellen sie dann für Recherchen zur Verfügung. Für die Suche auf Archie-Servern gibt es sowohl speziell entwickelte Software (Archie-Clients) als auch WWW-Gateways. Über Eingabemasken kann die Suche spezifiziert werden.
ein Archie-Gateways:
http://archie.th-darmstadt.de/
Eine wichtige Adresse ist FTPsearch in Trondheim (Norwegen): der Server liefert meistens blitzschnell Ergebnisse zurück:
http://ftpsearch.ntnu.no/ftpsearch/

Die bisherigen Hinweise gelten vor allem für die Suche nach frei kopierbarer Software (Public Domain, Freeware, Shareware usw.). Da kommerzielle Software nicht auf diesen Wegen im Internet vertrieben wird (Ausnahme: Demoversionen!), sollte man zuerst nach den Homepages der Hersteller bzw. Distributoren/Händler suchen.

Winfried Naumann

Literaturhinweise:

[1] Ulrich Babiak: Effektive Suche im Internet. Suchstrategien, Methoden, Quellen. (O'Reilly Essentials)
Köln: O'Reilly Verlag, 1997
[2] Ulrich Babiak's Homepage: http://www.dom.de/FreiRaum/uli/
[3] Franz Grieser: Webschnüffler. Internet Search Engines.
in: PC Professional 3/97, S. 208 ff.
(Suchmaschinen im Test: Alta Vista, Excite, Hotbot, Infoseek, Lycos und Yahoo!)
[4] Manfried Meyer: Wie Moses durchs Datenmeer. Kataloge und Indizes.
in: PC-Magazin DOS 6/97, S. 208 ff.
(Probleme, Hintergründe, Ausblick zu Katalogen und Indizes)
[5] Giesbert Damaschke: Wer suchet, der findet ... Manchmal ! Die richtigen Suchstrategien.
in: PC-Magazin DOS 6/97, S. 216 ff.
(spezielle Suchmöglichkeiten in den verschiedenen Katalogen und Indizes)