

5 Suchmaschinen, Robots und Agenten: Informationssuche im World Wide Web¹

Armin Günther & André Hahn

Ein entscheidender Nachteil der Architektur des Internet ist, daß die zu einem Zeitpunkt tatsächlich verfügbaren Informationen nirgends zentral verzeichnet werden. Nun wäre es recht aufwendig, um nicht zu sagen unmöglich, würde man die schätzungsweise 30 Millionen an das Internet angeschlossenen Computer (Stand: Januar 1998, Network Wizards, 1998) alle nach einer bestimmten Information durchsuchen, um dann unter Umständen trotzdem nicht fündig zu werden. Was Menschen unmöglich ist, erledigen spezielle Computer, die ebenfalls an das Internet angeschlossen sind. Suchmaschinen (Search Engines) durchkämmen laufend das Netz, sammeln soviel Informationen wie möglich auf und verzeichnen diese in Datenbanken, die dann von den Netzbewohnern mit Hilfe von Schlagwörtern durchsucht werden können. Andere Computer (z.B. *Archie*-Server) fungieren als Zentralen und übernehmen in regelmäßigen Abständen Bestandsdaten von anderen Rechnern. Und eine dritte Art von Zentralcomputer hat lediglich die Finger auf die Datenbestände anderer Computer ausgestreckt, die auf eine Anfrage hin unmerklich durchsucht werden (sog. *WAIS*-Server, Wide Area Information Server).

Ohne derartige Suchmaschinen würde das Internet erheblich an Wert verlieren, wäre der Nutzer doch darauf angewiesen, über bestehende, veränderte und neue Informationen im Netz unterrichtet zu werden. Nicht von ungefähr gehören Suchmaschinen zu den meistbesuchten Websites im Internet. Unter den „Top Ten“ der Web Domains, die das Unternehmen Media Metrix (Media Metrix, 1998) monatlich ermittelt, finden sich regelmäßig vier bis fünf Suchmaschinen (Yahoo!, Excite, Infoseek, Lycos und AltaVista). Doch auch der Umgang mit diesen Recherchewerkzeugen will gelernt sein. Welche Suchmaschinen gibt es überhaupt? Wie benutzt man sie? Wie kommen Suchergebnisse zustande und wie sind sie zu interpretieren? Im folgenden können

¹ Preprint vom 11. Juni 1998. Erscheint in Batinic, B. (1998) (Hrsg.). *Internet für Psychologen* (2te vollst. überarb. Auflage) (Kapitel 5). Göttingen: Hogrefe. Korrespondenzadressen: Dr. Armin Günther, Lehrstuhl für Ökonomische Psychologie, Universitätsstr. 16, D-86159 Augsburg, Telefon: (0821) 598 - 4089, Telefax: (0821) 598 - 4221, E-Mail: Armin.Guenther@wiso.uni-augsburg.de; Dipl.-Psych. André Hahn, Humboldt Universität zu Berlin, Lehrstuhl für Pädagogische Psychologie und Gesundheitspsychologie, Geschwister-Scholl-Str. 7, 10099 Berlin, Telefon: (+49-030) 2093-4024, FAX: (+49-030) 2093-4010, Email: hahn@psychologie.de, Web: <http://psychologie.de/hahn/>

diese Fragen zwar sicher nicht erschöpfend beantworten werden, doch soll zumindest soviel Basiswissen vermittelt werden, daß ein kompetenterer und effektiverer Umgang mit Suchmaschinen möglich wird. Behandelt wird im wesentlichen die Suche nach WWW-Dokumenten. Zusätzlich werden Hilfsmittel für die Suche nach E-Mail Adressen sowie nach Software vorgestellt.

5.1 Chaos 300 millionenfach: Ordnung und Orientierung im Web

Wieviele Dokumente das World Wide Web zum Abruf bereit hält, weiß niemand so genau. Klar ist nur, es sind sehr viele. Zwei anspruchsvolle wissenschaftliche Untersuchungen kamen kürzlich auf unterschiedliche, aber beachtliche Zahlen. Bharat und Broder (1998a, 1998b) schätzten die Größe des statischen, „indexierbaren Netzes“ im November 1997 auf über 200 Millionen und im März 1998 bereits auf mindestens 275 Millionen Textseiten. Lawrence und Giles (1998) kamen in ihrer Untersuchung für den Dezember 1997 sogar auf mindestens 320 Millionen Dokumente. Es liegt auf der Hand, daß eine Erschließung dieser Datenmenge nur mit Hilfe leistungsfähiger Suchwerkzeuge möglich ist.

Derartige Werkzeuge gibt es inzwischen in großer Zahl, ja das Angebot ist schon so groß und unübersichtlich geworden, daß bereits spezielle „Suchmaschinen-Suchmaschinen“ entwickelt wurden (vgl. 5.4). Die einzelnen Suchmaschinen unterscheiden sich in vielen Einzelheiten, so z. B. hinsichtlich der Anzahl und Auswahl der verzeichneten Dokumente, der Abfrageflexibilität, der Dauer einer Suche, der Ausführlichkeit der Rückmeldung und dem Ranking der Treffer. Bevor einzelne Suchsysteme vorgestellt werden, soll im folgenden zunächst eine Unterscheidung, die für das Verständnis von Suchmaschinen von besonderer Bedeutung ist, näher erläutert werden, die Unterscheidung zwischen (Link-)Verzeichnissen und Volltextsuchmaschinen (Indizes).

Linkverzeichnisse und Volltextsuchmaschinen

Suchmaschinen auf der Basis von (Link-)Verzeichnissen oder Katalogen sind im Grunde nichts anderes als große, thematisch gegliederte Sammlungen von (kommentierten) Verweisen zu WWW-Dokumenten. Sie entsprechen im Prinzip den Listen von Lesezeichen oder Bookmarks, die jeder WWW-Reisende mit Hilfe seines Browsers nach und nach ansammelt. Und so, wie man mit zunehmender Länge dieser Listen früher oder später beginnen wird, die gesammelten Links nach bestimmten Themen in „Ordner“ und „Unterordner“ zu sortieren, so sind auch die Links in verzeichnisbasierten Suchmaschinen in der Regel thematisch und hierarchisch organisiert. Beispielsweise findet man in *Yahoo!*, der wohl bekanntesten Suchmaschine dieser Art, Verweise zu psychologischen WWW-Dokumenten im Verzeichnis „Psychology“, das selbst im Verzeichnis „Social Science“ enthalten ist und seinerseits wiederum Unterverzeichnisse wie „Conferences“, „Counseling and Therapy“, „Organizations“ und „Tests and Experiments“ enthält. Im Unterschied zu den privaten Linksammlungen

sind die hier betrachteten Universalverzeichnisse für jeden im Internet frei zugänglich und umfassen Tausende von Verweisen zu allen möglichen Themen. Daneben gibt es natürlich auch themenspezifische Verzeichnisse, wie etwa den Online-Katalog von *psychologie.de*, der Links zu psychologischen WWW-Dokumenten enthält. Verzeichnisse, die sich speziell an Psychologinnen und Psychologen richten, werden im folgenden Kapitel behandelt (vgl. Hahn & Günther, in diesem Band). Hier geht es zunächst um allgemeine, nicht-fachspezifische Suchmaschinen.

Im Unterschied zu verzeichnisbasierten Suchmaschinen, die nur den Titel, die Kurzbeschreibung und die Internetadresse (URL) eines WWW-Dokumentes kennen, ist in *Volltextsuchmaschinen* der gesamte Text aller erfaßten Dokumente in Form von Wortlisten gespeichert. Eine Volltextsuchmaschine ist also gewissermaßen eine riesige Liste von Wörtern, bei der zu jedem Wort verzeichnet ist, in welchen WWW-Dokumenten es gefunden wurde. Sucht nun ein Anwender in diesem Index beispielsweise nach dem Wort „Psychologie“, so liefert die Suchmaschine alle Links von Dokumenten, in denen das Wort „Psychologie“ im Text enthalten war. Betrachten wir den Unterschied zwischen Linkverzeichnis und Volltextsuchmaschine an einem Beispiel: Angenommen, ein WWW-Dokument ist in einem *Linkverzeichnis* mit dem Eintrag: „Saisonal abhängige Depression: Eine Einführung in Diagnose und Therapie“ enthalten. Bei einer Suche in diesem Katalog würde das Dokument zwar bei dem Stichwort „Depression“ gefunden, nicht aber bei dem Stichwort „Lichttherapie“, auch dann nicht, wenn im Text des Dokumentes explizit von „Lichttherapie“ die Rede ist. Im Unterschied zu einem Linkverzeichnis kennt eine *Volltextsuchmaschine* den gesamten Text jedes indexierten WWW-Dokumentes und würde daher das Beispieldokument nicht nur bei einer Suche nach „Lichttherapie“ finden, sondern auch bei einer Suche nach Wörtern wie „regelmäßig“ oder „stark“, falls diese in dem entsprechenden Dokument vorkommen. Nur die in einer Sprache extrem häufig vorkommenden Wörter wie „die“, „ist“ oder „the“ werden auch von Volltextsuchmaschinen meistens nicht indexiert. Man spricht hier von „Stoppwörtern“.

Ein weiterer wichtiger Unterschied zwischen verzeichnisbasierten und Volltextsuchmaschinen besteht in der Art und Weise, wie diese generiert werden. Linkverzeichnisse werden normalerweise „manuell“, d. h. von Menschen, erstellt. Dazu sammeln Redakteure so viele Links wie möglich ein und ordnen diese den jeweils passenden Inhaltskategorien zu. Dieser Prozeß ist hochgradig selektiv: Webseiten von lediglich lokalem Interesse — wie das Mitgliederverzeichnis eines Vereins — oder auch untergeordnete Seiten, die durch eine Einstiegs- oder Homepage erreicht werden können, werden nicht eigens erfaßt. Im Unterschied hierzu können die Volltextsuchmaschinen zugrunde liegenden Indizes automatisch generiert werden, da hierzu kein inhaltliches Verständnis der Dokumente erforderlich ist. Spezielle Computerprogramme, Crawler, Robot oder Spider genannt, hangeln sich dazu von Link zu Link durch das World Wide Web, kopieren (fast) alle Dokumente, die sie finden und zerlegen diese in Wortlisten, gleichgültig, ob es sich um einen wissenschaftlichen Aufsatz, die aktuelle Speisekarte der Mensa oder ein Telefonverzeichnis handelt. Da Volltext-

suchmaschinen automatisch erstellt werden (aber auch, weil sie mehr oder weniger wahllos alles indexieren), verzeichnen sie erheblich mehr WWW-Dokumente als Suchmaschinen auf der Basis von Linkverzeichnissen. Letztere kommen bestenfalls auf einige Hunderttausend erfaßte Dokumente, während große Volltextsuchmaschinen wie AltaVista oder HotBot inzwischen die 100 Millionen Grenze überschritten haben (vgl. Tabelle 5.1).

Nach allen Unterschieden sei abschließend noch eine (unangenehme) Gemeinsamkeit von Linkverzeichnissen und Volltextsuchmaschinen genannt: Obwohl gute Linkverzeichnisse und Volltextsuchmaschinen ihre Datenbestände regelmäßig aktualisieren, geben sie doch immer nur ein Abbild des World Wide Webs, wie es einmal war. Es ist also keineswegs so, wie mancher Internet-Neuling meint, daß Suchmaschinen bei einer Anfrage blitzschnell den aktuellen Bestand des World Wide Webs durchsuchen. Anfragen werden vielmehr aus dem Archiv bedient. Der Nutzer wird daher öfter als im lieb ist die Erfahrung machen, daß das Dokument, auf das ein gefundener Link verweist, nicht mehr existiert, verschoben wurde oder seinen Inhalt verändert hat.

Was ist nun zur Recherche nach WWW-Dokumenten besser geeignet, die Suche im redaktionell erstellten Linkkatalog oder im automatisch generierte Volltextindex? Sowohl die Möglichkeit, nach Wörtern im gesamten Text eines WWW-Dokumentes zu suchen, als auch die wesentlich höhere Anzahl erfaßter Dokumente scheinen zunächst eindeutig für die Verwendung von Volltextsuchmaschinen und gegen (nicht-fachspezifische) verzeichnisbasierte Systeme zu sprechen. Hinzu kommt, daß gerade bei Universalverzeichnissen die wissenschaftlichen Angebote tendenziell immer mehr hinter dem Unterhaltungs- und Konsumbereich zurücktreten. So standen im Linkverzeichnis Yahoo! 854 Links in der Rubrik „Psychology“ 9.581 Links in der Rubrik „Movies and Film“ gegenüber (Stand: Mai 1998). Dabei muß man sich vor Augen halten, daß sich diese Kataloge nicht primär an Wissenschaftler oder Akademiker richten, sondern an ein breites Publikum. Da außerdem eine wesentliche Finanzierungsquelle von WWW-Katalogen in Werbeeinnahmen besteht, ist eine entsprechende „konsumfreundliche“ Selektion und Orientierung der Angebote vorprogrammiert. In der Tat spricht also viel für Volltextsuchmaschinen, wenn es um die Recherche nach Dokumenten zu bestimmten psychologischen Inhalten geht. Wer sich jedoch in einem Themen- oder Sachbereich erst einmal allgemein und eher ungezielt orientieren möchte oder wer eine Art Online-Branchenbuch sucht, für den kann eine verzeichnisbasierte Suchmaschine ein sinnvoller Einstieg sein.

Verzeichnisse und Volltextsuchmaschinen haben also ihre spezifischen Vor- und Nachteile, die man sehr vereinfacht auf die Formel „Klasse versus Masse“ bringen könnte. Die Tendenz geht daher auch unverkennbar in Richtung auf eine Kombination und Integration beider Ansätze. Allen weiter unten vorgestellten Volltextsuchmaschinen sind auch thematisch gegliederte Verzeichnisse angegliedert und umgekehrt übergeben thematische Linkverzeichnisse Anfragen bei Bedarf auch an Volltextsuchmaschinen. Wenn im folgenden Linkverzeichnisse und Volltextsuchmaschinen getrennt

behandelt werden, dann ist diese Trennung also eher analytisch zu verstehen: In der Praxis trifft der Nutzer in der Regel auf beide Systeme.

5.2 Verzeichnisbasierte Suchmaschinen: Internationale und deutsche Linkkataloge

Vom Blättern und Suchen

Verzeichnisbasierte Suchmaschinen kann man in der Regel auf zweierlei Art benutzen, die als *Blättern* („Browsing“) und als *Stichwortsuche* („Searching“) bezeichnet werden. Beim Blättern klickt man sich durch die Verzeichnishierarchie bis zu der relevanten Unterkategorie, um die dort versammelten Links durchzusehen. Auf diese Weise gelangt man beispielsweise bei einer Suche nach psychologischen Organisationen und Verbänden im Katalog von Yahoo! vom Verzeichnis „Social Science“ über das Verzeichnis „Psychology“ zum Verzeichnis „Organizations“. Bei der Stichwortsuche werden dagegen vom Suchenden frei wählbare Wörter oder Phrasen (z. B. „Therapie von Panikattacken“) in ein Formular eingegeben, nach denen dann alle oder ein Teil der verzeichneten Links und Verzeichnisnamen durchsucht werden. Alle Verweise und Verzeichnisse, bei denen die Suchwörter oder Phrasen im Titel oder im Kurzkomentar auftauchen, werden als Trefferliste ausgegeben. Mit einer solchen (globalen) Stichwortsuche unterläuft man also im Grunde die thematische Ordnung des Verzeichnisses, indem über alle Kategorien hinweg nach passenden Links gesucht wird. Das ist häufig sinnvoll, weil die Zuordnung von Verweisen zu Inhaltskategorien keineswegs immer eindeutig ist. Ja selbst die relevanten *Kategorien* zu finden, ist nicht immer ganz einfach, und auch hier kann die Stichwortsuche hilfreiche Dienste leisten. Wer würde schon die Rubrik „Psychology“ im Hauptverzeichnis „Good Life“, Unterverzeichnis „Hobbies & interests“ vermuten? Aber eben dort ist sie im Linkverzeichnis der Suchmaschine Infoseek zu finden!

Internationale Verzeichnisse

Im folgenden sollen zunächst einige Verzeichnisse etwas näher vorgestellt werden, die sich an ein internationales Publikum richten. Ein US-amerikanischer Schwerpunkt ist allerdings dennoch oft zu erkennen, was nicht verwundert, sind diese Kataloge doch alle in den USA zu Hause.

Yahoo! (<http://www.yahoo.com/>), seit Ende 1994 am Netz, ist nicht nur der Klassiker der verzeichnisbasierten Suchmaschinen, sondern außerdem eine der meistbesuchten Websites überhaupt (vgl. Media Metrix, 1998), eine Aktiengesellschaft, ein Markenprodukt und noch vieles mehr. Nach Angaben des Unternehmens umfaßt die Datenbank von Yahoo! über 500.000 Verweise, sortiert in mehr als 25.000 Inhaltskategorien (Yahoo!, 1998a). Diese Verweise können sowohl durch Blättern

(Browsing), als auch durch Stichwortsuche (Searching) erschlossen werden. Im zweiten Fall ist es möglich, die Suche auf die jeweilige Inhaltskategorie zu beschränken oder auf den gesamten Katalog auszudehnen. Wie bereits erwähnt sind in dem ausgezeichnet strukturierten hierarchischen Index auch einige hundert Links in der Kategorie „Psychology“ verzeichnet. Diese sind in weitere, sinnvolle thematische Untergruppen aufgeteilt. Zusätzliche psychologische Links lassen sich mit der Stichwortsuche finden, die sowohl innerhalb einer Kategorie, als auch über alle Kategorien hinweg durchgeführt werden kann. Wer im Linkverzeichnis von Yahoo! nicht fündig wird, dessen Suchanfrage wird an eine Volltextsuchmaschine übergeben. Dies ist gegenwärtig (Mai 1998) noch AltaVista, doch ist der Wechsel zu Inktomi angekündigt, einem Unternehmen, das unter anderem die Technologie von HotBot, einer AltaVista ebenbürtigen Volltextsuchmaschine, liefert.

Der guten Struktur von Yahoo! stehen allerdings deutliche Schwierigkeiten gegenüber, das Verzeichnis aktuell und relevant zu halten. Die Aufnahme neuer und die Aktualisierung alter Verweise ist oftmals außerordentlich langwierig und undurchsichtig, so daß manche Beobachter zu dem Schluß kommen, Yahoo! sei als Universalverzeichnis von WWW-Dokumenten an seine Grenzen gestoßen (Sullivan, 1998a). Dem entspricht, daß Yahoo! (wie die meisten anderen Suchmaschinen auch) sich immer mehr von einem Universalkatalog zu einem universalen Anbieter von Informationsdienstleistungen aller Art (z. B. Nachrichten-, Börsen-, Wetter-, Fernseh-, Sport-, Branchen-, Adressen-, Kommunikationsdienst) entwickelt. Neben diesen „Spezial-Yahooks“ gibt es inzwischen auch ein gutes Duzend „Länder-Yahooks“, eigenständige Linkverzeichnisse mit regionaler Ausrichtung. Hierzu gehört auch der Katalog Yahoo! Deutschland, der weiter unten vorgestellt wird.

Außer Yahoo! bieten noch eine ganze Reihe international bedeutsame, verzeichnisbasierte Suchmaschinen ihre Dienste an, von denen hier nur drei erwähnt werden sollen. Eines der jüngsten Verzeichnisse ist **Snap!** (<http://www.snap.com>), ein Service von c|net, dem „Computer Network“. Der Katalog ist sehr übersichtlich organisiert und erkennbar Yahoo! nachempfunden. Der Anspruch dieser Suchmaschine ist es, schnell, effektiv, und benutzerfreundlich hochwertige Suchergebnisse zu liefern. Der Akzent liegt also auf Qualität und weniger auf Quantität. Zumindest in der Rubrik „Psychology“ scheinen die Verweise in der Tat aktueller und relevanter zu sein als bei Yahoo!. Es bleibt abzuwarten, ob diese Aktualität auch auf die Dauer gewährleistet werden kann.

LookSmart (<http://www.looksmart.com>) verzeichnet nach eigenen Angaben 350.000 Websites in 20.000 Kategorien. Die Bedienung ist etwas umständlich, zumindest aber ungewohnt. So muß man beim Blättern (Browsing) immer erst die gesamte Inhaltshierarchie „durch klicken“, ehe man zu Verweisen gelangt, während Yahoo! oder Snap! bereits in den übergeordneten Kategorien Links zu Websites mit entsprechend übergreifender Thematik anbieten. LookSmart ist seit einiger Zeit auch den beiden größten Volltextsuchmaschinen, AltaVista und HotBot, als thematischer Katalog angegliedert.

Von eher historischem Interesse ist **Magellan** (<http://www.mckinley.com/>). In der Pionierphase des Internets war Magellan so etwas wie der Baedeker des World Wide Web. Ziel war weniger, ein Universalverzeichnis aufzubauen, sondern eher eine Art Ratingagentur einzurichten, die die Qualität von Websites beurteilt. Zu über 60.000 Websites wurden Kurzbesprechungen angefertigt und die entsprechenden Angebote mit maximal vier Sternen bewertet. Nach dem Vorbild von Magellan gibt es heute zahlreiche, mehr oder weniger aussagekräftige Auszeichnungen von Websites. Die Qualifikation als „Magellan 4-Star Site“ galt jedenfalls noch bis vor Kurzem als ein begehrtes Gütesiegel. Gegenwärtig (Mai 1998) ist der Reviewprozeß ausgesetzt und es ist offen, ob, wann und in welcher Form er wieder aufgenommen wird. Magellan wird inzwischen von Excite Inc. betrieben, einem Unternehmen, das vor allem durch seine Volltextsuchmaschine bekannt ist.

Deutsche Verzeichnisse

Während die bisher vorgestellten Verzeichnisse einen internationalen Anspruch haben, gibt es auch Kataloge, die regional beschränkt und konsequenterweise auch in der jeweiligen Landessprache verfaßt sind. Erfäht und geordnet wird hier in erster Linie das WWW-Angebot einer bestimmten Region. Im folgenden sollen sechs Kataloge kurz beschrieben werden, die das WWW-Angebot im deutschsprachigen Raum erschließen.

Yahoo! Deutschland (<http://www.yahoo.de>) ist einer von über zehn „Länder-Yahos“ und versteht sich als Verzeichnis der wichtigsten deutschsprachigen Websites im WWW. Das in München lokalisierte Unternehmen arbeitet redaktionell unabhängig von der amerikanischen Muttergesellschaft. Auch bei Yahoo! Deutschland wird der Katalog durch eine Reihe spezieller, auf die Bedürfnisse des deutschsprachigen Publikums abgestimmte Informationsdienste wie Nachrichten- und Finanzdienste, Wettermeldungen und Branchenbuch ergänzt.

Web.de (<http://www.web.de>) bezeichnet sich selbst als „Deutschlands umfangreichster Web-Katalog“. Gegenwärtig (Stand: Mai 1998) sind knapp 130.000 deutschsprachige Webseiten verzeichnet. Die Navigation durch das übersichtliche Verzeichnis erfolgt wahlweise durch Blättern oder Suchen, wobei sich eine Suche auf jeden Fall empfiehlt, um relevante Kategorien zu finden. „Psychologie“ beispielsweise wird hier unter den „Geisteswissenschaften“ geführt, doch weitere Verweise kann man unter anderem in den Kategorien „Organisationen: Medizin: Kliniken: Psychotherapie“, „Wissenschaft: Medizin: Fachbereiche: Psychotherapie“ oder „Wirtschaft: Branchen: Medizin: Alternative Medizin: Praxen“ entdecken. Auch dieser Katalog wird durch zusätzliche Informationsdienste wie Nachrichten, einem Routenplaner oder Surftips ergänzt.

Dino (<http://www.dino-online.de>), das **Deutsche InterNet Organisationssystem**, ist seit 1995 online und ebenfalls auf deutschsprachige WWW-Seiten spezialisiert. Die über 120.000 Verweise sind in 2.500 Rubriken sortiert, die durchblättert und durchsucht werden können. Verglichen mit Web.de ist die Darstellung der Links und Ru-

briken etwas unübersichtlicher. Natürlich dürfen auch hier die Zusatzangebote nicht fehlen, über die sich Webkataloge mehr und mehr zu profilieren suchen. Dino bietet den Nutzern einen eigenen Newskanal mit den (bebilderten) Nachrichten von Reuters, ein Branchenbuch, Surftips und einen Chat Bereich.

Zwei weitere, eher kleine deutsche WWW-Kataloge sind Allesklar und Sharelook. **Allesklar** (<http://www.allesklar.de/>) verzeichnet ca. 80.000 Adressen aus Deutschland, Österreich und der Schweiz, sowie deutschsprachige Seiten anderer Länder, sortiert in 2.000 Kategorien (Stand: März 1998). Neben der Standardoberfläche, die der von LookSmart entspricht (einschließlich der dort beschriebenen Probleme), wird auch eine dem Microsoft-Explorer nachempfundene Java-Version der Suchmaschine angeboten. Spezialität von Allesklar ist eine leistungsfähige Regionalsuche, bei der Stichwörter und Orte frei kombiniert werden können oder das WWW-Angebote von 25 vorgegebenen Regionen durchblättert werden kann.

Auch bei **Sharelook** Deutschland (<http://www.sharelook.de/>) kann das WWW-Angebot einiger weniger deutscher Städte durchblättert (nicht aber durchsucht) werden. Der Gesamtkatalog umfaßt nach Angaben des Betreibers 75.000 Einträge in rund 10.500 Rubriken (Stand: Mai 1998). Neben Sharelook Deutschland gibt es auch auf Österreich (<http://www.sharelook.at/>), die Schweiz (<http://www.sharelook.ch/>) und Frankreich (<http://www.sharelook.fr/>) ausgerichtete Kataloge.

Gerade für eine geographisch orientierte Suche in Deutschland ist **Entry.de** (<http://www.entry.de/>) sehr nützlich. Dieser Service des DFN-Vereins listet nach Städten geordnet WWW-Server in Deutschland. Private Homepages findet man hier zwar nicht, wohl aber die Startseiten von Bildungseinrichtungen, Internet-Providern, Firmen, öffentlichen Einrichtungen, Vereinen, Gesellschaften und Parteien. Das Blättern in diesem werbefreien Katalog ist einfach und problemlos, die Ausgabe übersichtlich.

5.3 Volltextsuchmaschinen

Die Recherche in einer Volltextsuchmaschine erfolgt ausschließlich über die Stichwortsuche, das für verzeichnisbasierte Suchmaschinen charakteristische „Blättern“ durch eine Themenhierarchie entfällt. Verglichen mit verzeichnisbasierten Suchmaschinen ist die effektive Nutzung von Volltextsuchmaschinen wesentlich anspruchsvoller. Geht es dort hauptsächlich darum, in der jeweiligen Themenhierarchie die relevanten Kategorien ausfindig zu machen, so besteht das zentrale Problem bei Volltextsuchmaschinen darin, aus vielen Millionen Dokumenten eine Handvoll relevanter Texte heraus zu filtern, indem man ein paar Worte oder Phrasen angibt, die in den gesuchten Texten enthalten sein könnten. Dabei unterstützt einen keine thematische Vorsortierung der Dokumente, kein Redakteur, der vorweg einen Qualitätscheck durchgeführt hätte und keine Kurzbeschreibungen des jeweiligen Inhalts eines WWW-Dokuments. Während selbst ein so großes Verzeichnis wie Yahoo! zum Stichwort

„psychology“ „nur“ 1050 Links (und 67 Kategorien) findet, sind es bei der Volltextsuchmaschine AltaVista über 700.000 Treffer — 700.000 Dokumente also, die irgendwo das Wort „psychology“ enthalten (Stand: Mai 1998). Um derartige Suchinstrumente sinnvoll einsetzen zu können, ist es daher unerlässlich, sich mit den jeweiligen Möglichkeiten, die Suche nach Dokumenten einzugrenzen, vertraut zu machen. Doch diese Mühe lohnt sich, denn verglichen mit redaktionell erstellten Linkverzeichnissen erschließen Volltextsuchmaschinen das World Wide Web nicht nur sehr viel umfassender, sondern meist auch bedeutend aktueller. Gerade beim Auffinden von wissenschaftlichen Informationen leisten diese Rechercheinstrumente unschätzbare Dienste.

Spieglein, Spieglein an der Wand, wer ist die Größte im ganzen Land?

Zu den wesentlichen Qualitätsmerkmalen einer Volltextsuchmaschine gehört zweifellos die Anzahl der indexierten Dokumente — wenngleich Größe, wie wir noch sehen werden, nicht alles ist. Die Entwickler von AltaVista, der wohl bekanntesten Volltextsuchmaschine, verfolgten dann auch das ehrgeizige Ziel, möglichst alle WWW-Dokumente, die zu einem gegebenen Zeitpunkt verfügbar sind, zu erfassen und zu indexieren (AltaVista, 1998a). Wie nahe die verschiedenen Suchmaschinen diesem Ziel kommen, ist nur schwer festzustellen, unter anderem deshalb, weil man hierzu eine verlässliche Schätzung der Gesamtzahl aller WWW-Dokumente benötigt. Zu den vermutlich sorgfältigsten Untersuchungen, die die Netzabdeckung verschiedener großer Suchmaschinen zum Gegenstand hatten, gehören die oben bereits genannten Studien von Bharat und Broder (1998a, 1998b) sowie Lawrence und Giles (1998). Vor allem die im renommierten Wissenschaftsmagazin Science veröffentlichte Untersuchung von Lawrence und Giles machte Schlagzeile mit dem Ergebnis, daß die einzelnen Suchmaschinen einen geringeren Anteil der verfügbaren WWW-Dokumente erfassen, als bisher angenommen wurde (vgl. Tabelle 5.1).

Tabelle 5.1: Anzahl indexierter WWW-Seiten in ausgewählten Suchmaschinen

	L & G (12/1997)	B & B (3/1998)	Betreiberinfo. (5/1998)	de Seiten (5/1998)
AltaVista	91 Mio.	110 Mio.	140 Mio.	6 Mio.
HotBot	110 Mio.	100 Mio.	110 Mio.	5 Mio.
Nothern Light	65 Mio.	—	—	—
Excite	45 Mio.	33 Mio.	50 Mio.	—
Infoseek	32 Mio.	33 Mio.	—	3 Mio.
Lycos	10 Mio.	—	30 Mio.	—
geschätzte Größe des WWW	320 Mio.	275 Mio.	—	—

Erläuterungen. **L & G** = Lawrence & Giles (1998), **B & B** = Bharat & Broder (1998b), **Betreiberinfo.** = im Web verfügbare Angaben der Suchmaschinen-Betreiber, **de Seiten** = Anzahl der Treffer bei eine Suche nach WWW Seiten der Domain de

Am besten schnitten in allen Fällen die Suchmaschinen *HotBot* und *AltaVista* ab, wenn auch in unterschiedlicher Reihenfolge: Nach Lawrence und Giles erfaßt HotBot etwa jede Dritte der geschätzten 320 Millionen WWW-Seiten, gefolgt von AltaVista. In den Untersuchungen von Bharat und Broder — zwei Mitarbeitern von NEC, dem Unternehmen also, das auch AltaVista betreibt! — war die Reihenfolge der beiden Suchmaschinen umgekehrt: AltaVista kam im März 1998 auf 40% und HotBot auf 36% (wobei allerdings das gesamte indexierbare WWW auf „nur“ 275 Millionen Dokumente geschätzt wurde). Inzwischen haben die Betreiber von AltaVista erklärt, noch einmal 40 Millionen Dokumente zusätzlich erfaßt zu haben und nun mit einer Größe von 140 Millionen indexierten Webseiten unangefochten die Nummer Eins zu sein (AltaVista, 1998b). Wie auch immer die genauen Zahlen aussehen mögen: Der Abstand der beiden Spitzenreiter AltaVista und HotBot zu den ebenfalls sehr populären Suchmaschinen Excite, Infoseek und Lycos ist jedenfalls zur Zeit eindeutig und unbestritten. Den Anschluß an die Spitze kann allenfalls Nothern Light halten, eine neuere Suchmaschine, die insbesondere eine interessante Methode bietet, die gefundenen Dokumente in Kategorien zu organisieren.

Relevanzranking und Abfrageflexibilität

Doch Größe ist nicht alles. Die Leistungsfähigkeit einer Volltextsuchmaschine hängt wesentlich auch von ihrer Retrievalsoftware ab, also von jenen Programmen, mit deren Hilfe der Nutzer den angesammelten Datenmassen die für ihn relevanten Informationen zu entlocken sucht. Eine Suchmaschine, die das World Wide Web vollständig indexieren würde, wäre praktisch wertlos, könnte man die erfaßten Dokumente immer nur nach einem einzigen Stichwort gleichzeitig durchsuchen und würden die Treffer in willkürlicher Ordnung ausgegeben. Zu fast jedem Stichwort erhielte man auf diese Weise hunderte, tausende oder gar hunderttausende Treffer, was die Verarbeitungskapazität jedes Nutzers hoffnungslos überfordern würde. Je größer der Datenpool, desto intelligenter muß offensichtlich auch die Retrievalsoftware sein — und desto intelligenter muß der Suchende seine Abfrage formulieren. Auf der anderen Seite nutzt natürlich auch die intelligenteste Retrievalsoftware nichts, wenn die relevanten Informationen nicht in der Datenbank enthalten sind. Je spezifischer oder seltener die gesuchte Information im Netz ist, desto wichtiger wird offenbar eine möglichst umfassende Erfassung der WWW-Dokumente.

Die Retrievalsoftware von Volltextsuchmaschinen unterstützt die zielgerichtete Suche im wesentlichen auf zweierlei Weise: Erstens indem sie den Nutzer in die Lage versetzt, die Dokumente, die er sucht, möglichst genau zu beschreiben und zweitens durch ein zusätzliches Relevanzranking aller Dokumente, auf die diese Beschreibung zutrifft. Die gefundenen Treffer werden dann nach Relevanz geordnet ausgegeben.

Sehen wir uns zunächst kurz den zweiten Punkt, das Relevanzrating, an. Alle Volltextsuchmaschinen versuchen abzuschätzen, wie zentral die vom Nutzer vorgegebenen Suchworte oder Begriffe in den gefundenen Dokumenten sind: Kommt das Suchwort nur beiläufig im Text vor oder bezeichnet es das zentrale Thema dieses Do-

kumentes? Da Suchmaschinen die indexierten Texte inhaltlich nicht verstehen, sind sie auf formale Kriterien zur Abschätzung der Relevanz eines Dokumentes angewiesen. Zu einer Erhöhung des Relevanzratings kommt es beispielsweise dann, wenn Suchwörter oder -phrasen im Titel einer Seite enthalten sind, wenn sie auf einer Seite gehäuft vorkommen, wenn mehrere der vorgegebenen Suchausdrücke im Text enthalten sind, wenn sie nahe beieinander vorkommen usw. Neben der Berücksichtigung von Platzierung, Verteilung, Häufigkeit und Konfiguration von Suchwörtern greifen einige Suchmaschinen auch noch auf Kontextwissen zurück, etwa darüber, wonach üblicherweise mit einem bestimmten Wort gesucht wird. In der Regel haben die *Benutzer* einer Suchmaschine keine Möglichkeit, dieses Relevanzrating selbst zu beeinflussen und zu konfigurieren. Es gibt allerdings Ausnahmen: In der Expertensuche von AltaVista können Wörter als Rankingkriterien vorgegeben werden und bei Lycos kann die Gewichtung verschiedener Kriterien eingestellt werden. Darüber hinaus haben die *Autoren* von WWW-Seiten durch die Einfügung von sogenannten Meta-Tags in das Dokument die Möglichkeit, der Suchmaschine zentrale Schlagwörter und Themen der jeweiligen Seite mitzuteilen. Nicht alle Suchmaschinen nutzen diese Informationen allerdings, und manche WWW-Autoren mißbrauchen diese Möglichkeit, um (meist aus kommerziellem Interesse) eine möglichst gute Platzierung bei möglichst vielen populären Suchwörtern wie „sex“ oder „erotic“ zu erzielen, auch wenn es im Dokument um ganz etwas anderes geht (spamming).

Während der Nutzer einer Volltextsuchmaschine auf das Relevanzranking in der Regel also keinen Einfluß nehmen kann, stehen ihm verschiedene Möglichkeiten zur Verfügung, die Treffermenge durch geschickt formulierte Suchanweisungen einzugrenzen und dadurch die Relevanz der Treffer insgesamt zu erhöhen. Diese Möglichkeiten scheinen allerdings von vielen Nutzern nicht genutzt zu werden. Sie begnügen sich mit der Eingabe von ein oder zwei Suchwörtern, in der Hoffnung, die Suchmaschine werde ihnen schon die passenden Informationen liefern (Infoseek, 1998a). Karzauninkat (1998) vergleicht eine solche „Suchstrategie“ mit dem Vorgehen eines Bibliotheksbesuchers, der auf die Frage des Bibliothekars, was er wünsche, auch nur mit einem einzigen Wort — z. B. „Spaghetti“ — antwortet:

„Können Sie sich also vorstellen, in welcher Lage eine noch so ausgetüftelte Suchmaschine steckt, wenn Sie ihr einen einzelnen, einsamen Begriff zuwerfen? Sie wird sich bitter rächen, indem sie Ihnen alles, was sie irgendwie zum Thema Spaghetti findet, zurückwirft. Zack. Auf einen großen Haufen. 134.000 Treffer“ (Karzauninkat, 1998, Kapitel: Bedienung der Suchmaschinen).

Obwohl Volltextsuchmaschinen ihr Bestes tun, selbst noch bei derart spärlichen Informationen die „besten Treffer“ zuerst auszugeben — Seiten, in denen „Spaghetti“ eine tragende Rolle spielen —, ist doch klar, daß ohne weitere Selektionskriterien und Kontextinformationen keine gezielte Suche möglich ist. Bei einem Umfang von bis zu 140 Millionen Dokumenten wird fast jedes Wort in Hunderten oder gar Tausenden von Texten enthalten sein, und bei der Eingabe von nur einem Wort hat auch die klüg-

ste Volltextsuchmaschine kaum eine Chance, ein brauchbares Relevanzranking vorzunehmen. Die Hilfsseiten von Volltextsuchmaschinen — die man sich im übrigen unbedingt ansehen sollte! — betonen daher auch immer wieder, daß durch die Angabe von mehreren Wörtern die Zahl irrelevanter Treffer erheblich gesenkt wird. Zumindest aber erhält die Suchmaschine zusätzliche Hinweise, welche Verweise an den Anfang der Liste zu setzen sind (nämlich die, bei denen möglichst viele Suchwörter möglichst häufig an möglichst zentralen Stellen im entsprechenden Dokument zu finden sind). Klar dürfte auch sein, daß seltene und spezifische Wörter häufigen und allgemeinen vorzuziehen sind. Manchmal sind auch ungewöhnliche Personennamen hilfreich. Flowforscher haben es in dieser Beziehung leicht, denn mit „Csikszentmihalyi“ läßt sich die Treffermenge bereits erheblich eingrenzen, wenngleich AltaVista selbst diesen Namen noch in mehr als 1.700 Dokumenten findet! Statt also von Suchmaschine zu Suchmaschine zu springen und diesen immer wieder nur „Spaghetti“ — oder „Psychologie“ — zuzurufen, ist es deshalb in der Regel sinnvoller, sich etwas intensiver mit den Suchoptionen und der Syntax der ausgewählten Suchmaschine zu befassen, um auf diese Weise gezieltere Anfragen formulieren zu können. Im folgenden werden einige wenige Konzepte und Hilfsmittel erläutert, die sich bei vielen Volltextsuchmaschinen finden und die sinnvoll eingesetzt die Recherche erheblich erleichtern können.

Das Einmaleins der Volltextsuche

Die Möglichkeiten, differenzierte und komplexe Suchabfragen zu konstruieren, unterscheiden sich von Suchmaschine zu Suchmaschine. Tabelle 5.2 zeigt für ausgewählte Systeme, welche (syntaktischen) Möglichkeiten dem Nutzer jeweils offen stehen, durch die Formulierung seiner Abfrage die Treffermenge zu bestimmen. Die verschiedenen Konzepte werden im folgenden erläutert.

Tabelle 5.2: Suchoptionen ausgewählter Suchmaschinen

	Phr	G/K	Trnk	Operatoren	Selektoren	globale Selektionen
AltaVista	ja	ja	ja	<i>simple search</i> : +, - <i>advanced search</i> : and, or, not, near, (...)	anchor, applet, domain, host, image, link, text, title, url	Sprache, Datum
HotBot	ja	ja	nein	+, -, AND, OR, NOT, (...),	domain, depth, feature (acrobat, applet, activex, audio, embed, flash, form, frame image, script, shockwave, table, video, vrm), linkdomain, linkext, scriptlanguage, title, after, before, within	—

Nothern Light	ja	nein	nein	+, -, and, or, not, (...)	url, title, text	—
Excite	ja	nein	nein	+, -, AND, OR, AND NOT, (...)	—	—
Infoseek	ja	ja	nein	+, -,	link, site, title, url	—
Lycos	ja	nein	auto	+, -, and, or, not, adj/n, oadj/n, near/n, onear/n, far/n, ofar/n, before	—	URL, Titel, Domain
Deutsche Volltextsuchmaschinen						
Fireball	wie AltaVista					
Inter-search	ja	nein	ja	+, -, and, or, not, (...), near, [...], ?	title, meta, url, email	—
Eule	nein	nein	nein	and, or	—	—

Erläuterungen. **Phr** = Phrasensuche, **G/K** = unterscheidet zwischen Groß- und Kleinschreibung (case sensitivity), **globale Selektionen**: Sucheingrenzung, die nicht Bestandteil eines Suchausdrucks sein kann.

Wörter und Phrasen. Alle ausgewählten Volltextsuchmaschinen geben dem Nutzer die Möglichkeit, nicht nur nach einzelnen Wörtern, sondern auch nach zusammenhängenden Wortfolgen oder „Phrasen“ zu suchen. Häufig werden Phrasen durch Anführungszeichen abgegrenzt, in anderen Fällen kann man über ein Formular einstellen, ob die Suchwörter als Phrase oder als Liste von Einzelwörtern verstanden werden sollen. Mit der Eingabe von

„internet für psychologen“

läßt sich beispielsweise bei AltaVista nach der entsprechenden Wortfolge suchen.

Groß- und Kleinschreibung. Als Faustregel kann gelten: Klein schreiben! Bei Kleinschreibung suchen die gebräuchlichen Suchmaschinen *sowohl klein- als auch* großgeschriebene Varianten der Suchausdrücke, während bei Großschreibung je nach Suchmaschine ganz unterschiedliches, manchmal merkwürdiges, geschehen kann. Gerade für deutschsprachige Anwender ist in dieser Beziehung die Suchmaschine Infoseek besonders heimtückisch. Diese interpretiert eine Folge von zwei großgeschriebenen Wörtern als Personennamen (wie „Peter Müller“) und führt *automatisch* eine Phrasensuche durch. Wer also bei Infoseek mit der Eingabe von

Bulimie Psychotherapie

nach Informationen zu Psychotherapie bei Bulimie sucht, wird gegenwärtig durch die Meldung „No Results Found“ überrascht. Grund: Infoseek sucht erfolglos nach der *Phrase* „Bulimie Psychotherapie“. Gibt man dagegen

bulimie psychotherapie

ein, so findet Infoseek knapp 3.000 Seiten, da nun nach „Bulimie“ und/oder „Psychotherapie“ unabhängig von der Groß- oder Kleinschreibung gesucht wird. Fazit: Großschreibung sollte man nur *gezielt* einsetzen, d. h. nur dann, wenn man weiß, welchen Effekt sie bei der jeweiligen Suchmaschine hat und wenn man eben diesen Effekt erzielen will.

Trunkierung und Platzhalter. Manche Suchmaschinen (z. B. AltaVista) erlauben bei Bedarf auch die Suche nach Wortbestandteilen, manche (z. B. Lycos) suchen Wortbestandteile immer mit und wieder andere (z. B. HotBot) suchen immer nur nach vollständigen Wörtern. In AltaVista kann man beispielsweise durch die Eingabe von `psychotherap*`

nach „Psychotherapie“, „Psychotherapien“, „Psychotherapeutin“, „psychotherapy“, „psychotherapies“ usw. suchen. Gelegentlich, wie bei der deutschen Suchmaschine Intersearch (und wohl auch bei AltaVista), kann man auch *innerhalb* eines Wortes beliebige Buchstaben durch Platzhalter ersetzen (z. B. „autor*en“ für „Autoren“, „Autorinnen“, „autorisieren“, „Autorennen“ usw.).

Operatoren. Alle der im folgenden betrachteten Volltextsuchmaschinen stellen die +/- Syntax zur Verfügung, die eine einfache und effektive Möglichkeit bietet, eine Suche zu begrenzen. Ein Pluszeichen vor einem Wort oder einer Phrase gibt an, daß der entsprechende Ausdruck in einem Dokument *enthalten sein muß*, ein Minuszeichen, daß er *nicht enthalten sein darf*. (Ausnahme: Bei Infoseek führt das Minuszeichen nicht zum Ausschluß aus der Treffermenge, sondern zu einer Abwertung im Relevanzranking, was aber in der Praxis zumindest bei großer Trefferzahl, auf dasselbe hinaus läuft). Die Eingabe von

`+bulimie +psychotherapie`

reduzierte dann auch die Trefferliste bei Infoseek von knapp 3.000 Seiten, in denen mindestens eines der beiden Wörter enthalten ist, auf 23 Seiten, in denen beide Wörter enthalten sind. Durch die Eingabe von

`+memory cognition -RAM`

weist man AltaVista (in der Einfachsuche) an, nach Seiten zu suchen, die „memory“ enthalten müssen, „cognition“ enthalten sollten und „RAM“ nicht enthalten dürfen. (Die Großschreibung von „RAM“ bewirkt bei AltaVista, daß nur Seiten entfernt werden, in denen der Ausdruck in genau dieser Schreibweise enthalten ist.).

Komplexere Suchabfragen lassen sich häufig mit sogenannten *Boole'schen Operatoren* bilden. Mit Ausnahme von Infoseek bieten alle im folgenden vorgestellten Suchmaschinen dazu die Operatoren AND, OR und NOT an (die teilweise großgeschrieben werden *müssen*, teilweise großgeschrieben werden *können*). Gelegentlich kommen noch weitere Operatoren wie z. B. NEAR oder ADJ (für „adjacent“, also „benachbart“) hinzu, bei denen es sich strenggenommen nicht um Boole'sche Opera-

toren handelt. Mit Hilfe von Klammern lassen sich die logischen Beziehungen festlegen, z. B.

```
(mood OR emotion) AND (memory OR cognition)
```

Selektoren und globale Selektionen. Einige Suchmaschinen bieten die Möglichkeit, gezielt nach Wörtern oder Phrasen *in bestimmten Bereichen* von WWW-Dokumenten zu suchen, beispielsweise innerhalb der Titel, innerhalb der Links oder innerhalb der URLs von WWW-Seiten. Mittels spezieller Schlüsselwörter, den Selektoren (Eingabe meist in der Form Selektor:Suchausdruck) oder über die Auswahl in einem Formular legt man den Bereich oder das Feld fest, in dem gesucht werden soll. Eine solche „Feldsuche“ kann unter Umständen äußerst nützlich sein. In der Expertensuche bei AltaVista ist es beispielsweise möglich, durch die Eingabe von

```
coping AND (domain:de OR domain:at OR domain:ch)
```

die Suche nach dem Stichwort „coping“ auf Server in Deutschland, Österreich und der Schweiz zu beschränken. Gibt man

```
link:http://www.meinrechner.de/meinehomepage.html
```

ein, so sucht AltaVista alle Seiten heraus, die den definierten Link enthalten (von denen also auf meine Homepage verwiesen wird). Stellt man bei der Suchmaschine HotBot zunächst die Option „Boolean phrase“ ein und sucht dann nach

```
psychophysics AND (feature:video OR feature:audio)
```

so werden Seiten nachgewiesen, die das Wort „psychophysic“ enthalten und in die außerdem Video- oder Audio-Dateien eingebunden sind. Häufiger anwendbar ist vermutlich die Suche nach einem Wort oder einer Phrase im *Titel* eines WWW-Dokumentes. Gibt man etwa in der Standardsuche von AltaVista

```
+title:"klinische psychologie" studienordnung
```

ein, werden Seiten gesucht, bei denen die Phrase „klinische psychologie“ im Titel vorkommen muß und bei denen außerdem irgendwo (also nicht unbedingt im Titel) auch noch das Wort „studienordnung“ enthalten sein sollte.

Bei vielen Suchmaschinen werden derartige Feldsuchen durch Formulare erheblich erleichtert, in denen man die jeweiligen Suchfelder per Mausklick auswählen kann, ohne sich um die korrekte Syntax kümmern zu müssen. In Kauf nimmt man dafür allerdings gelegentlich eine gewisse Einschränkung an Flexibilität. Ähnlich wie Feldsuchen funktionieren *globale Selektionen*. Der Unterschied zur Feldsuche besteht darin, daß diese Selektionskriterien nicht als einzelne Suchworte in eine komplexe Abfrage integriert werden können, sondern ein notwendiges Kriterium für alle Treffer darstellen. So ist bei AltaVista die Datumseingrenzung nur als globale Selektion möglich, bei HotBot dagegen als Feldsuche.

Internationale Volltextsuchmaschinen

Im folgenden sollen einige Volltextsuchmaschinen porträtiert werden, die für die Suche nach WWW-Dokumenten zur Verfügung stehen.

AltaVista (<http://www.altavista.digital.com/>) verzeichnet nach Angabe des Betreibers inzwischen 140 Millionen WWW-Einträge (Stand: Mai 1998) und nimmt damit für sich den Spitzenplatz unter den Volltextsuchmaschinen zumindest in puncto Größe in Anspruch (vgl. Tabelle 5.1). Neben der Suche nach WWW-Dokumenten können übrigens auch die aktuellen Beiträge aller Newsgruppen nach Stichworten durchsucht werden. AltaVista bietet sowohl eine Standard-, als auch eine erweiterte Suchoption, wobei eine jeweils unterschiedliche Syntax zu beachten ist. Zu beachten ist bei der erweiterten Suche auch, daß die für das Ergebnisranking relevanten Wörter eigens eingegeben werden müssen bzw. können — sonst erfolgt die Ausgabe unsortiert. Mit der Möglichkeit, unterschiedliche Feldsuchen in komplexe Boole'sche Ausdrücke (oder auch in die einfachere +/- Syntax) zu integrieren, erlaubt AltaVista die Formulierung sehr differenzierter Suchausdrücke über den vermutlich umfangreichsten Volltextindex von WWW-Dokumenten, der gegenwärtig verfügbar ist. Damit kommt AltaVista speziell für wissenschaftliche Recherchen zweifellose eine Spitzenposition unter den Volltextsuchmaschinen zu.

Eine gerade auch für deutschsprachige Nutzer interessante Spezialität von AltaVista ist die automatische Spracherkennung und die Möglichkeit, eine Suche auf Dokumente in einer bestimmten Sprache zu begrenzen. Das ist besonders dann hilfreich, wenn man beispielsweise deutschsprachige Texte nach Stichwörtern durchsuchen möchte, die wie „coping“, „attribution“ oder auch „Psychologie“ in verschiedenen Sprachen vorkommen. Im Unterschied zu anderen Volltextsuchmaschinen, die allenfalls eine Begrenzung auf einen bestimmten Sprachraum bieten, indem sie die Suche auf die entsprechenden Top-Domains begrenzen (z. B. auf die Domains de, at, ch, und li für den deutschsprachigen Raum), bietet AltaVista zusätzlich eine genuine Spracherkennung, die auf einer Analyse der erfaßten Texte selbst basiert. Kombiniert mit der Beschränkung auf bestimmte Domains, kann man in AltaVista beispielsweise problemlos nach englischsprachigen Webseiten auf „deutschen“ Servern (domain:de) oder nach deutschsprachigen Seiten an amerikanischen Universitäten und anderen Bildungseinrichtungen (domain:edu) suchen.

Für deutschsprachige Nutzer von AltaVista gibt es auch eine deutschsprachige Benutzeroberfläche. Dazu muß man auf den nordeuropäischen Server altavista.telia.com wechseln. Die Suchergebnisse auf dem internationalen (amerikanischen) und dem nordeuropäischen Server sind im allgemeinen allerdings nicht identisch. Aktueller ist in jedem Fall der internationale Server. Neuerdings bietet AltaVista auch die Möglichkeit, WWW-Dokumente automatisch übersetzen zu lassen. Die Resultate von Babelfish, dem Übersetzungsprogramm, sind häufig äußerst amüsant, können aber zumindest einen ungefähren Eindruck davon geben, worum es auf einer Seite geht. Neben der Beschränkung auf eine bestimmte Sprache, läßt sich in der erweiterten Suchoption die Recherche global auch auf WWW-Dokumente aus einem bestimmten Zeitraum

(Datum der letzten Bearbeitung bzw. Aufnahme in die Datenbank) begrenzen. Das ist etwa dann nützlich, wenn man nach neueren oder veränderten WWW-Seiten sucht.

HotBot (<http://www.hotbot.com>) ist gegenwärtig offenbar die einzige in puncto Größe (Anzahl der indexierten Dokumente) und Abfrageflexibilität AltaVista ebenbürtige Suchmaschine. Zumindest bei Abfragen mittlerer Komplexität ist die Benutzerfreundlichkeit von HotBot der von AltaVista vermutlich sogar überlegen, da man hier die Suchlogik und die meisten weiteren Einstellungen bequem über ein WWW-Formular auswählen kann (aber nicht muß). Wie AltaVista, so bietet auch HotBot verschiedene Möglichkeiten für Feldsuchen. In Verbindung mit Boole'schen Ausdrücken können hiermit komplexe Suchabfragen formuliert werden. Sogar die Datumseingrenzung kann differenziert in Boole'sche Ausdrücke integriert werden. Folgender komplexer Ausdruck, mit dem nach der Phrase „visual perception“ (obligatorisch) und dem Wort „psychology“ (fakultativ) innerhalb von Dokumenten gesucht wird, die während des letzten Monats verändert oder neu aufgenommen wurden, ist beispielsweise in der „Boolean phrase“ Eingabe von HotBot gültig:

```
(+"visual perception" psychology) AND within:1/months
```

Ausgestattet mit diesen Merkmalen ist HotBot eine Suchmaschine, die sowohl dem „Normalverbraucher“ als auch dem Suchexperten hervorragende Möglichkeiten für die Recherche nach WWW-Dokumenten bietet.

Was HotBot im Unterschied zu AltaVista vor allem fehlt, ist die Spracherkennung. Eine Einschränkung der Suche auf deutschsprachige Dokumente ist hier also nicht möglich, wohl aber eine Einschränkung auf Domains (z. B. de, au, ch, edu). Eine ebenfalls wichtige Restriktion ist, daß HotBot im Unterschied zu AltaVista nicht die Möglichkeit zur Trunkierung von Wörtern bietet.

Nothern Light (<http://www.nlsearch.com/>) ist der Shooting-Star unter den Volltextsuchmaschinen. Erst seit August 1997 am Netz, hat sie sich bereits zu einer der führenden Suchsysteme entwickelt. Interessant ist Northern Light nicht nur wegen ihrer umfangreichen Datenbank (vgl. Tabelle 5.1), sondern auch wegen einer neuartigen Organisation der Suchergebnisse und dem ungewöhnlichen Unternehmenskonzept. Bei der Ausgabe der Suchergebnisse wird der Versuch gemacht, die Orientierungsvorteile eines Linkverzeichnisses mit der Suchtiefe einer Volltextsuchmaschine zu kombinieren. Die Ergebnisse werden dazu nicht nur, wie üblich, einem Relevanzranking unterzogen und in einer einzigen langen Liste ausgegeben, sondern zugleich auch einer automatischen hierarchischen Klassifikation unterzogen und in dynamisch generierte Ordner („Custom Search Folders“) sortiert. Für die Zusammenfassung in einem gemeinsamen Ordner gibt es vier Kriterien und entsprechend vier Typen von Ordnern: gleiches Thema (z. B. Depression, Fußball, Expertensysteme), gleicher Dokumententyp (z. B. Presseerklärungen, Produktbeschreibungen, Landkarten), gleiche Quelle (kommerzieller Anbieter, Ausbildungsinstitution) oder gleiche Sprache (z. B. Englisch, Deutsch, Französisch). Öffnet man einen Ordner, so werden die enthaltenen Verweise möglicherweise abermals geordnet. Allzuviel Intelligenz darf man von dieser automatischen Klassifikation allerdings nicht erwarten und inwieweit sich dadurch die

Trefferquote einer Suche erhöht, ist schwer zu sagen. Auf jeden Fall erlaubt dieses Verfahren alternative Zugriffe auf ein Suchergebnis, womit die Chance gegeben ist, auch Dokumente aufzuspüren, die bei einer einzigen langen Ergebnisliste eine hintere Platzierung bekommen.

Neuartig an Nothern Light ist auch das Unternehmenskonzept. Nicht Werbung, sondern der Verkauf von Information soll die finanzielle Basis bereitstellen. Zusätzlich zu der kostenlosen Netzrecherche werden dazu auch die Texte von derzeit etwa 2 Millionen Aufsätzen aus 3.400 Zeitschriften, darunter auch einige psychologische und sozialwissenschaftliche Fachzeitschriften, zum Kauf angeboten. Die Suche nach diesen Aufsätzen ist in die Suchmaschine integriert, Titel und Kurzzusammenfassung gibt es kostenlos und erst beim Erwerb des Volltextes wird eine Gebühr fällig. Diese fällt bei einem Standardpreis von \$1 (Stand: Mai 1998) ungewöhnlich günstig aus.

Die Betreiber von **Excite** (<http://www.excite.com>) heben als Besonderheit ihrer Suchmaschine hervor, daß diese nicht nur nach Wörtern, sondern auch nach „Konzepten“ sucht. Basis dieser „Intelligent Concept Extraction“ (ICE) genannten Technologie ist eine Analyse von Wortkonfigurationen oder -clustern innerhalb der indextierten Dokumente. Auf dieser Grundlage hält Excite bei einer Abfrage auch nach Seiten Ausschau, die Wörter enthalten, die im allgemeinen mit den Suchwörtern zusammen gebraucht werden. Dies führt dazu, daß in der Trefferliste auch Seiten enthalten sein können, in denen die Suchwörter überhaupt nicht auftauchen. Ähnlich wie etwa die „Refine“ Funktion von AltaVista bietet Excite auf dieser Grundlage auch Hilfen bei der Verfeinerung der Suchabfrage, indem zusätzliche Suchwörter vorgeschlagen werden, die mit den vom Nutzer eingegebenen Suchwörtern zusammenhängen. Auf Wunsch kann die Ergebnisliste nach Websites geordnet werden. Grob gesagt bedeutet dies, daß WWW-Seiten von demselben Server zusammen erscheinen, was wesentlich zur Übersichtlichkeit beitragen kann. Nützlich ist auch die für jeden Treffer angebotene Option, nach ähnlichen Seiten zu suchen. Mit der alternativ zur Standardsuche angebotenen „Power search“ betreibt Excite in gewisser Weise Etikettenschwindel, da hier lediglich die in der Standardsuche über die +/-Syntax zur Verfügung stehenden Möglichkeiten (umständlicher) über ein Formular generiert werden. Das Unternehmen betreibt neben der Volltextsuchmaschine auch den kleineren Index **Webcrawler** (<http://www.webcrawler.com>), sowie einige regionale Excites, darunter **Excite Deutschland** (<http://www.excite.de/>). Sieht man von der deutschsprachigen Benutzeroberfläche, den auf deutsche Nutzer ausgerichteten zusätzlichen Informationsdiensten, eventuellen Geschwindigkeitsvorteilen und natürlich auch der deutschen Werbung ab, so bringt Excite Deutschland für die eigentliche Recherche nach WWW-Dokumenten selbst keine zusätzlichen Vorteile, da sie keine eigene Datenbank unterhält, sondern lediglich die Suche im internationalen Excite-Index auf die Top-Domain de begrenzt.

InfoSeek (<http://www.infoseek.com/>) präsentiert auf seiner Homepage die Volltextsuchmaschine und das eigene Linkverzeichnis Seite an Seite und dokumentiert so die Gleichrangigkeit der beiden Suchsysteme. Tatsächlich behauptet das Unternehmen,

über das größte Linkverzeichnis im Web zu verfügen, insbesondere also auch noch Yahoo! zu übertreffen. Was die Größe des Volltextindexes angeht, so gehört Infoseek mit Excite und Lycos (s. u.) nur zur zweiten Liga. Dafür kann die Volltextsuchmaschine durch eine einfache und funktionale Bedienung überzeugen, die allerdings die Flexibilität von AltaVista oder HotBot nicht erreicht. So läßt Infoseek als einzige der hier vorgestellten Suchmaschinen keine Boole'schen Ausdrücke zu. Dafür sind immerhin vier verschiedene Feldsuchen möglich, und das erweiterte Suchformular stellt eine komfortable Möglichkeit dar, diese Optionen zu nutzen, ohne sich mit der Syntax der Suchmaschine herumzuschlagen. Die Ergebnisliste bietet einen guten Kompromiß zwischen Übersichtlichkeit (durch Sortierung nach Websites und Ausgabe nur der jeweils „relevantesten“ Seite einer Site) und Vollständigkeit (durch die Möglichkeit, die Gruppierung nach Sites wieder aufzuheben). Sehr einfach und problemlos ist auch die Einschränkung einer Suche auf die Treffer einer vorhergehenden Suche. Relevante Kategorien des Linkverzeichnisses werden ebenfalls bei einer Suche im Volltextindex als Hyperlink ausgegeben, so daß mit einem Mausklick die Suche im thematischen Katalog von Infoseek fortgesetzt werden kann. Wie Excite, so versucht sich auch Infoseek durch eine besonders intelligente Retrievalsoftware zu profilieren. Im Mai 1998 informierte das Unternehmen über die Einführung einer neuen Suchtechnologie unter dem vielversprechenden Namen „Extra Search Precision“ (ESP), welche bei der Suche nach populären Themen zu „dramatisch“ verbesserten Suchresultaten“ bei der Eingabe von nur ein oder zwei Suchworten (wie „airplanes“ oder „car rentals“ oder „stocks“) führe (Infoseek, 1998a). Da diese Technologie aber in erster Linie auf die Unterstützung von Ein- bis Zweiwort-Abfragen zu populären Themen zielt, ist der Nutzen von ESP für Recherchen nach wissenschaftlichen Spezialthemen eher fraglich.

Auch Infoseek bietet eine deutsche Benutzeroberfläche und die Möglichkeit, die Suche auf „Deutschland“ zu beschränken. Doch wie bei Excite bedeutet dies nur, daß die Suche im Infosseek-Katalog auf die Domain de begrenzt wird.

Lycos (<http://www.lycos.com/>), seit 1994 online, ist eine der ältesten Volltextsuchmaschinen im Netz, die 1996 durch ein Linkverzeichnis ergänzt wurde. Ähnlich wie bei Yahoo!, sind bei Lycos heute diverse Informationsdienste eng verwoben mit Konsum- und Unterhaltungsangeboten. Selbst wenn man annimmt, daß die Ergebnisse von Lawrence und Giles in diesem Fall deutlich zu niedrig ausgefallen sind, so ist die Netzabdeckung der Volltextsuchmaschine eher gering, weit entfernt von den Werten, die AltaVista und HotBot erreichen. Dennoch zieht Lycos mehr Nutzer an als diese, was nicht zuletzt an der Plazierung dieser Suchmaschine in den Suchseiten der Internet Browser von Netscape und Microsoft liegen dürfte. Lycos bietet eine Standard- und eine Fortgeschrittenensuche mit einer Reihe von Optionen. Ungewöhnlich ist, daß die Gewichtung der Kriterien für das Trefferranking individuell konfiguriert werden kann. Ähnlich wie bei Excite und Infoseek ist auch bei Lycos die Trefferliste nach „Sites“ sortiert, mit den oben erwähnten Orientierungsvorteilen. Und wie bei Excite kann auch hier zu jedem Treffer nach ähnlichen Seiten gesucht werden („More Like This“). Man vermißt allerdings eine Datumseingrenzung und auch die Trunkierung

von Suchwörtern scheint nicht (mehr) möglich zu sein. Die in der Fortgeschrittenensuche angebotenen Eingrenzung auf eine bestimmte Sprache entpuppt sich bei genauerem Hinsehen als Domainbegrenzung: Die Wahl von Deutsch (bzw. German) als Sprache beispielsweise ist tatsächlich eine Beschränkung der Suche auf die Domains de, au, ch und li. Eine eigentliche Spracherkennung, wie bei AltaVista, findet also nicht statt. Wie wir es schon bei Excite und Infoseek gesehen haben, so bietet auch Lycos Deutschland (<http://www.lycos.de/>) für die WWW-Recherche keinen eigenen Index an, sondern nur die Beschränkung auf Domains in deutschen Sprachraum (de, at, ch, li).

Auch Lycos bietet neben der Volltextsuche einen eigenen Linkkatalog an. Durch den Kauf des Contentproviders WiseWire hat sich das Unternehmen ein Verzeichnis („Web Guides“) zugelegt, das im Unterschied zu herkömmlichen Katalogen weitgehend automatisch generiert wird. Das zeigt sich beispielsweise negativ darin, daß es zu den einzelnen Websites keine Kurzbeschreibungen gibt. Den Input für die einzelnen Kategorien liefert eine intelligente Suchmaschine, das Relevanzrating innerhalb der Kategorien wird dann von den Nutzern selbst vorgenommen, indem diese über die „Qualität“ der einzelnen Sites abstimmen können („Do you like it?“). Das Unternehmen erhofft sich dadurch nicht zuletzt auch eine stärkere Bindung der Nutzer an „ihr“ Verzeichnis.

Deutsche Volltextsuchmaschinen

Was „deutsche Suchmaschinen“ sind, ist nicht ohne weiteres klar. Mehrere Interpretationen sind denkbar: Deutsche Suchmaschinen sind (1) Suchmaschinen, die innerhalb des deutschsprachigen Raums entwickelt wurden, (2) Suchmaschinen, die auf Computern innerhalb Deutschlands oder des deutschsprachigen Raums betrieben werden, (3) Suchmaschinen, die deutschsprachige WWW-Dokumente nachweisen, (4) Suchmaschinen, die Dokumente auf Computern im deutschsprachigen Raum (oder die Dokumente innerhalb der Domains de, at, ch und li nachweisen), (5) Suchmaschinen mit deutschsprachiger Benutzeroberfläche, (6) Suchmaschinen mit deutscher Werbung. Tatsächlich gibt es hier keine verbindliche Definition. Man hat aber mitunter den Eindruck, daß die Einführung neuer, nationaler Suchmaschinen hauptsächlich dazu dient, Unternehmen neue Werbeplattformen zu erschließen, die eine gezieltere Ansprache (beispielsweise) deutscher Konsumenten ermöglichen. Denn bei vielen Suchmaschinen ist es fraglich, ob sie dem Anwender tatsächlich irgend einen Zusatznutzen gegenüber vorhandenen Systemen bringen. Das gilt weniger für nationale *Linkverzeichnisse*, die wegen ihrer Selektivität und der Kommentierung der Links in der jeweiligen Landessprache Sinn machen. Nationale *Volltextsuchmaschinen* bringen dagegen nur dann einen Vorteil, wenn sie die WWW-Dokumente der jeweiligen Region umfassender, aktueller, differenzierter oder schneller erschließen, als es internationale Systeme ohnehin schon tun. (Von den erweiterten Informationsdiensten wird hier abgesehen.) Oft ist jedoch das Gegenteil der Fall: Die nationalen Volltextsuchmaschi-

nen erschließen auch das „Nation Wide Web“ häufig schlechter als die großen internationalen Systeme.

So dürfte gegenwärtig keine „deutsche“ Volltextsuchmaschine eine vergleichbare Anzahl von deutschsprachigen oder im deutschen Sprachraum beheimateten WWW-Dokumenten indexiert haben, wie AltaVista oder HotBot (vgl. Tabelle 5.1). Entsprechende Eingrenzungen der Suche auf deutschsprachige Domains (z. B. durch die Eingabe von +domain:de in der Standardsuche von AltaVista) oder auf deutschsprachige Dokumente (bei AltaVista) sind leicht möglich. Auch bei Infoseek und Lycos sind entsprechende Einschränkungen der Top-Domains verfügbar, weshalb sich ein Besuch auf den deutschen Zweigstellen dieser Suchmaschinen unter Umständen erübrigt. Nur bei Excite muß man zu Excite Deutschland wechseln, um eine auf deutschsprachige Domains begrenzte Suche durchzuführen. Zusätzlich zu den oben genannten nationalen Niederlassungen internationaler Volltextsuchmaschinen, sollen daher im folgenden nur noch drei regionale Volltextsuchmaschinen vorgestellt werden, die deutschsprachige Dokumente oder Dokumente im deutschsprachigen Raumnachweisen.

Fireball (<http://www.fireball.de/>) hieß zunächst „Flipper“, dann „Kitty“ und ist sicher eine der besten genuin „deutschen“ Volltextsuchmaschinen für deutschsprachige Internetseiten. Immerhin 4,6 Millionen Dokumente sollen es gegenwärtig (Stand: Mai 1998; Kalb, 1998) sein, die mit Hilfe der Datenbank dieser Suchmaschine nachgewiesen werden können, womit Fireball vermutlich die einzige deutsche Volltextsuchmaschine ist, die das deutschsprachige Web in etwa so umfassend erfaßt, wie AltaVista oder HotBot. Indexiert werden nur deutschsprachige Seiten, vorausgesetzt das Spracherkennungsprogramm identifiziert diese zutreffend. Die Software, die die Anfragen bearbeitete, stammt von AltaVista. Für Nutzer dieser amerikanischen Suchmaschine hat das den Vorteil, daß sie für ihre Suchanfragen bei Fireball die vertraute und sehr leistungsfähige Syntax verwenden können. Recherchen nach Dokumenten in anderen Sprachen können von der Fireball-Homepage an AltaVista weitergereicht werden. Wie viele andere Suchmaschinen, so reichert auch Fireball seine Basisfunktion durch zusätzliche Dienstleistungen wie Surf tips oder Life Suche (zusehen, wonach andere gerade suchen) an. Mehr als eine bloße Zugabe ist „Paperball“, eine Suchmaschine, die die aktuellen Online-Ausgaben von über 40 deutschsprachigen Tageszeitungen durchsucht.

Intersearch (<http://www.intersearch.de/>) ist ebenfalls eine deutschsprachige Volltextsuchmaschine mit ca. 3,2 Millionen indexierten Seiten (Stand: Juni 1998; Monz, 1998). Im Index enthalten sind deutschsprachige WWW-Dokumente, aber auch nicht-deutschsprachige Dokumente aus den Domains de und au. Es kann zwischen einer Suche in Deutschland und einer Suche in Österreich unterschieden werden. Daneben bietet Intersearch eine ganze Reihe von Möglichkeiten, komplexe Abfragen zu formulieren. Unterstützt wird nicht nur, wie üblich, die Phrasen-Suche und die +/- Syntax, sondern auch Boole'sche und eine Reihe weitere Operatoren, Trunkierung (und Wildcards), sowie vier Möglichkeiten der Feldsuche: Suche in den URLs, den sog. Meta-

Tags, den Titeln und E-Mail-Links von WWW-Dokumenten. Interessant ist insbesondere die Möglichkeit, E-Mail-Links auf Webseiten zu durchsuchen.

Eule (<http://www.eule.de/>) verwaltet mit etwa 2 Millionen WWW-Dokumenten (Stand: Juni 1998; Schoebel, 1998) einen beachtlichen Datensatz, der allerdings deutlich hinter Fireball zurückbleibt. Wie bei Fireball ist der Index auf deutschsprachige Dokumente beschränkt, die Domain spielt für die Aufnahme keine Rolle (Schoebel, 1998). Die Abfrageflexibilität dieser Suchmaschine ist gering und erreicht bei weitem nicht die Möglichkeiten von Fireball oder Intersearch. Zur Verfügung steht lediglich die übliche +/- Syntax. An Boole'schen Operatoren werden zwar „and“ und „or“ (alternativ „und“ und „oder“) unterstützt, da jedoch keine Klammern verwendet werden können, ist damit keine Erweiterung der Abfragemöglichkeiten verbunden. Als einzige hier vorgestellte Suchmaschine erlaubt Eule keine Phrasensuche.

5.4 Meta-Tools und Spezialsuchmaschinen

So positiv die Existenz einer Vielzahl von Suchmaschinen auf den ersten Blick erscheinen mag, sie macht das Suchen nach Informationen im Netz nicht eben einfacher. Denn die verschiedenen Suchmaschinen decken keineswegs alle denselben Bereich des Internets ab, womit die Wahl eines solchen Werkzeugs letztlich nur eine Frage der Funktionalität und Erreichbarkeit wäre. Die unterschiedlichen Suchmaschinen „kennen“ vielmehr auch unterschiedliche WWW-Seiten, was zur Folge hat, daß mit unterschiedlichen Suchmaschinen durchaus auch Unterschiedliches gefunden werden kann. Bharat und Broder (1998a) etwa kommen in ihrer Studie zu dem Schluß, daß von der Gesamtheit der Seiten, die von den vier Suchmaschinen AltaVista, HotBot, Excite und Infoseek zusammengenommen erfaßt werden, weniger als 1,4% von *allen* vier Suchmaschinen indexiert werden — eine überraschend kleine Schnittmenge! Lawrence und Giles (1998, S. 100) fanden umgekehrt, daß durch die Kombination der Ergebnisse aller sechs in die Untersuchung einbezogenen Suchmaschinen etwa 3,5 mal so viele Dokumente erfaßt wurden, wie durch eine einzelne Maschine. Bei der Verwendung von nur zwei Suchmaschinen, erzielte die Kombination von HotBot und AltaVista die größte Netzabdeckung. Hinzu kommt, daß sich die Suchmaschinen auch in der Art des Relevanzrankings unterscheiden, was unterschiedliche Dokumente an den Anfang der Trefferliste bringen kann. Wer also wirklich gründlich recherchieren will (oder muß), der kommt manchmal nicht umhin, mehrere Suchmaschinen zu verwenden.

Zur Unterstützung einer solchen extensiven Suche haben einige WWW-Entwickler Webseiten erstellt, die die Suche über mehrere Suchmaschinen hinweg unterstützen sollen (vgl. Tabelle 5.3). Im einfachsten Fall handelt es sich hierbei um *Verzeichnisse von Suchmaschinen*, die in der Regel nach irgend einer Systematik geordnet sind. Etwas anspruchsvoller sind Verzeichnisse, die für die aufgeführte Suchmaschine gleich auch ein Eingabefeld bereit stellen, von dem aus eine Anfrage an die jeweilige

Search Engine geschickt werden kann. Die Ergebnisse erhält der Nutzer dann direkt von der jeweiligen Suchmaschine. Diese Seiten versammeln also gewissermaßen die Eingangstüren zu diversen Suchmaschinen an einem Platz., und entsprechend nennt Karzauninkat (1998) eine solche Website einen „*Sammelsucher*“. Noch weiter gehen die „eigentlichen“ *Meta-Suchmaschinen*. In ihrer anspruchsvollsten Variante wird hier von *einem* Eingabefeld aus eine Anfrage an verschiedene Suchmaschinen geschickt, die Suchergebnisse werden von der Meta-Suchmaschine eingesammelt, Doubletten heraus gefiltert, die verbleibenden Treffer sortiert und dem Anfrager die resultierende Liste übermittelt.

Tabelle 5.3: Suchmaschinenverzeichnisse und Meta-Suchmaschinen

Titel	World Wide Web Adresse (URL, http://)
Suchmaschinenverzeichnisse (Auswahl)	
Beaucoup	www.beaucoup.com/
Metaindex Europe	www.hj.se/hs/bibl/miewww/
Sammelsucher (Auswahl)	
All-In-One	www.albany.net/allinone/
Edison! Search	www.edison.com/search.htm
eDirectory	www.edirectory.com/
Klug-Suchen	www.klug-suchen.de/
SEARCH.COM	www.search.com/
The Internet Sleuth	www.isleuth.com/
CUSI der TU Chemnitz	www.tu-chemnitz.de/misc/cusi.html
CUSI der Uni. Düsseldorf	www.biophys.uni-duesseldorf.de/gateways/cusi.html
CUSI der Uni. Freiburg	www.informatik.uni-freiburg.de/~mannes/
CUSI der Uni. Siegen	www.unix-ag.uni-siegen.de/search/
CUSI light der Uni. Siegen	www.unix-ag.uni-siegen.de/search/quick.html
OneClick (Cusi)	www.psychologie.de/internet/oneclick.html
Meta-Suchmaschinen (Auswahl)	
MetaCrawler	www.metacrawler.com/
Savvy Search	guaraldi.cs.colostate.edu:2000/
ProFusion	profusion.ittc.ukans.edu/
MetaGer	meta.rrzn.uni-hannover.de/
Ahoy! The Homepage Finder	ahoy.cs.washington.edu:6060/

Verzeichnisse von Suchmaschinen gibt es in großer Zahl. Nicht zuletzt die oben vorgestellten allgemeinen Linkverzeichnisse enthalten in der Regel auch eine Rubrik mit Suchmaschinen. Bei Yahoo! etwa findet man ein Verzeichnis „Search Engines“ unter „Computers and Internet: Internet: World Wide Web: Searching the Web“. Ein Beispiel für ein spezielles Suchmaschinenverzeichnis mit Links zu über 1.000 Suchmaschinen, Linkverzeichnissen und Online-Datenbanken ist **Beaucoup**. Auf den Nachweis europäischer Suchmaschinen hat sich der **Metaindex Europe** spezialisiert, in dem die Links noch Ländern sortiert sind.

Auch *Sammelsucher* findet man zahlreich im Netz. Technisch ist es kein großes Problem, die Eingabefelder von Suchmaschinen auf einer WWW-Seite zu vereinen, zumal die an möglichst vielen Besuchern interessierten Suchmaschinen selbst eigene Anleitungen zu einem solchen „Web-Napping“ veröffentlichen. Ein klassisches Beispiel für einen umfangreichen Sammel-sucher ist **All-In-One**. Alleine für die Recherche nach WWW-Dokumenten werden hier die Eingabefelder von über 70 Suchmaschinen versammelt, hinzu kommen noch zahlreiche Suchmaschinen mit anderen oder spezifischen Funktionen (z. B. für die Suche nach Software, Personen, Publikationen). Ein anderes Beispiel ist das Verzeichnis von **Edison! Search**, das von sich behauptet, die größte Sammlung von Suchinstrumenten im Internet zu bieten. Solche Superlative sind allerdings mit Vorsicht zu genießen, da derartige Verzeichnisse dazu tendieren, jede beliebige Online-Datenbank als Suchmaschine aufzulisten. Ähnlich wie im oben erwähnten Metaindex Europe sind die Suchmaschinen auch im **eDirectory** nach Ländern geordnet, nur diesmal mit weltweiter Abdeckung und natürlich dem für Sammel-sucher charakteristischen Eingabefeld zu jeder Maschine. Manche Sammel-sucher können nicht nur durchblättert, sondern auch ihrerseits durchsucht werden. In diesem Fall kann der Nutzer mittels frei wählbarer Stichworte nach Spezialsuchmaschinen zu einem bestimmten Thema (z. B. Gesundheit) oder nach Suchmaschinen einer bestimmten Art (z. B. Suchmaschinen für Newsgruppen) suchen. Im deutschsprachigen Raum ist **Klug-Suchen** ein Beispiel für einen solchen durchsuchbaren Sammel-sucher. Große internationale Verzeichnisse von Suchmaschinen, die durchblättert und durchsucht werden können, sind **SEARCH.COM** und **The Internet Sleuth**, letzteres mit über 3.000 Einträgen. Diese Sammel-sucher sind vor allem sinnvoll um Spezial-Suchmaschinen und Online-Datenbanken mit Informationen zu spezifischen Themen zu finden und den ersten Zugriff auf diese erleichtern. Die intensivere Recherche kann dann vor Ort, also bei dem jeweils ausgewählten Suchsystem, erfolgen.

Komfortabler noch in der Bedienung sind „**CUSIs**“ (Configurable Unified Search Engine). Bei diesen Sammel-suchern hat nicht jede aufgeführte Suchmaschine ihr eigenes Suchfeld, vielmehr gibt man den Suchausdruck in ein Eingabefeld ein, wählt dann eine Suchmaschine in einem Drop-Down Menü aus und startet die Suche. Für Nutzer aus dem deutschsprachigen Raum besonders geeignet sind CUSIs an den Universitäten Chemnitz, Düsseldorf, Freiburg und Siegen. Besonders für den Einsteiger empfehlenswert ist die CUSI aus Siegen, da hier jede auswählbare Suchmaschine noch kurz beschrieben wird. Darüber hinaus können erfahrene Sucher in Siegen auch auf eine kompakte „CUSI light“ Version zurückgreifen. Unter dem Dach von „psychologie.de“ bietet auch einer der Autoren dieses Kapitels, André Hahn, eine umfangreiche CUSI unter dem Titel „**OneClick**“ an.

Teils Suchmaschinenverzeichnis, teils Sammel-sucher sind im übrigen auch die Suchseiten, zu denen ein Klick auf das „Suchen“-Menü der WWW-Browser von Netscape oder Microsoft führt. Welche Suchmaschinen wie auf diesen Suchseiten erscheinen, oder auch wie oft eine Suchmaschine beim Laden dieser Seiten als Standard vorausgewählt ist, ist nicht zuletzt eine Frage des Geldes (Sullivan, 1998b). Die Un-

ternehmen, die die großen Suchmaschinen betreiben, zahlen viele Millionen Dollar für eine (möglichst gute) Plazierung auf den Suchseiten von Netscape oder Microsoft. Über die auf diese Weise erzielten wesentlich höheren Besucherzahlen, erhoffen sich die Betreiber dieser Suchmaschinen entsprechend höhere Werbeeinnahmen.

Kommen wir zu den „eigentlichen“ Meta-Suchmaschinen. Der oben bereits erwähnte „**Internet Sleuth**“ beispielsweise ist zwar in erster Linie ein „Sammelsucher“, besitzt zusätzlich aber bereits Funktionen einer „echten“ Meta-Suchmaschine. So kann man von hier aus eine Suchabfrage an mehrere Volltextsuchmaschinen und Linkverzeichnisse gleichzeitig abschicken. Die Ergebnisse werden vom „Internet Detektiv“ (Sleuth = Detektiv) eingesammelt und in einer gemeinsamen Liste ausgegeben. Allerdings findet keine Integration der verschiedenen Trefferlisten statt, vielmehr werden die jeweils ersten Ausgabeseiten der kontaktierten Suchmaschinen einfach aneinander gehängt.

Im Unterschied hierzu führt beispielsweise **MetaCrawler** eine vollständige Integration der verschiedenen Suchergebnisse durch. Diese Meta-Suchmaschine schickt die jeweilige Anfrage an Lycos, Infoseek, WebCrawler, Excite, AltaVista und Yahoo!. Von jeder dieser Suchsysteme werden maximal die ersten 30 Treffer eingesammelt, in ein einheitliches Format gebracht, Duplikate entfernt bzw. zusammengefaßt, einem erneuten Ranking unterzogen (das sich aus der Summe der normalisierten Rankings der einzelnen Suchmaschinen ergibt) und an den Nutzer ausgegeben. Die Abfrageflexibilität bei Meta-Suchmaschinen ist naturgemäß eher gering, da im Grunde nur diejenigen Suchoptionen zugelassen werden können, die von *allen* abzufragenden Suchmaschinen unterstützt werden. Manche Meta-Suchmaschinen machen hier allerdings Kompromisse, indem sie eine Abfrage den Möglichkeiten der jeweiligen Suchmaschine anpassen. Bei MetaCrawler stehen dem Nutzer wahlweise eine Oder- und eine Und-Verknüpfung der Suchwörter zur Verfügung oder auch eine Phrasensuche. Darüber hinaus kann auch die +/- Syntax verwendet werden und die Suche auf Dokumente bestimmter Domains bzw. Domaingruppen begrenzt werden.

Eine integrierte Ergebnisliste liefern auch **Savvy Search** und **ProFusion**. Der Nutzer kann hier teilweise selbst auswählen, welche Suchmaschinen in die Meta-Suche mit einbezogen werden sollen. Weitere internationale Meta-Suchmaschinen sind bei Karzauninkat (1998), Rabinovitch (1997) oder Sullivan (1998c) beschrieben.

In und vor allem *für* Deutschland gibt es zwar nur eine Meta-Suchmaschine, diese ist dafür aber besonders gut gelungen, so daß für weitere Meta-Suchmaschinen hierzulande kaum Bedarf ist. Mit **MetaGer** — ein Service des Regionalen Rechenzentrums für Niedersachsen und des Lehrgebiets Rechnernetze und Verteilte Systeme an der Universität Hannover — können 11 deutsche Linkverzeichnisse oder Volltextsuchmaschinen und zusätzlich noch die deutschsprachigen Dokumente von AltaVista gleichzeitig durchsucht werden. Damit nicht genug, läßt sich auch noch MetaCrawler als internationale Meta-Suchmaschine zuschalten. MetaGer liefert eine integrierte Ergebnisliste und verfügt darüber hinaus noch über eine Reihe attraktiver Merkmale. So können die in die Suche einzubeziehenden Suchdienste frei gewählt und Suchzeit und

Treffermenge eingestellt werden. Auf Verlangen überprüft die Suchmaschine sogar die gefundenen Webseiten auf ihre Existenz und gibt sie nach ihrem Änderungsdatum geordnet aus — ein Service, der verständlicherweise mit erheblich längeren Wartezeiten einher geht. Kleine Details, wie der Umstand, daß MetaGer während einer Suche alle 10 Sekunden meldet, von welcher Suchmaschine die Ergebnisse noch ausstehen, eine übersichtliche Trefferstatistik oder die Suche nach „Quick Tips“ in einer lokalen Datenbank runden den vorzüglichen Eindruck ab, den diese Meta-Suchmaschine hinterläßt.

Das Grundkonzept von Meta-Suchmaschinen, die Datenbestände der verschiedenen Suchmaschinen gewissermaßen zu integrieren, ist intuitiv überzeugend. Es stellt sich die Frage, ob damit die Nutzung einzelner Suchmaschinen nicht überflüssig ist? Die Frage ist aus mehreren Gründen zu verneinen: Erstens ist eine Vergrößerung des Suchraums häufig gar nicht erforderlich. Das Problem einer Internet-Recherche besteht ja im Gegenteil häufig in einem *Übermaß* an *irrelevanten* Treffern. Woran es mangelt, sind also nicht so sehr größere als vielmehr intelligentere Suchmaschinen. „Intelligenter“ als ihre Zulieferer sind die üblichen Meta-Suchmaschinen aber sicher nicht, eher umgekehrt: Zumindest ihre Abfrageflexibilität ist notgedrungen reduziert. Und dies ist ein zweiter Grund, warum Meta-Suchmaschinen keine grundsätzliche Alternative zur Einzelabfrage von Suchmaschinen bieten. Wie oben bereits erwähnt, kann eine Meta-Suchmaschinen uneingeschränkt nur jene Suchoptionen bieten, die *alle* angeschlossenen Suchmaschinen unterstützen. Überspitzt gesagt kann sie also höchstens so „intelligent“ sein, wie die „dümme“ ihrer Komponenten. Es macht wenig Sinn, in einer Meta-Suchmaschine eine komplexe Anfrage zu formulieren, die von den meisten der angeschlossenen Suchmaschinen nicht verstanden wird. Ein weiteres, weniger schwerwiegendes Problem ist, daß Meta-Suchmaschinen auch nicht so schnell sein können, wie eine einzelne der datenliefernden Suchmaschinen. Das gilt besonders dann, wenn eine integrierte Ergebnisliste ausgegeben werden soll. Hier bestimmt die langsamste der durchsuchten Datenbanken das Tempo. Andererseits ist eine Meta-Suchmaschine natürlich erheblich schneller als es die sukzessive Abfrage der einzelnen Suchmaschinen wäre. Kurz: Die Suche in einer Meta-Suchmaschine ist sicher eine sinnvolle Ergänzung, aber kein Ersatz für die intensive Suche in einer einzelnen Suchmaschine.

Möglicherweise werden aus diesen Gründen zukünftig verstärkt intelligentere, statt umfassendere Such- und Meta-Suchmaschinen entwickelt. Ein gutes Beispiel hierfür ist **Ahoy!**. Die Intelligenz dieser Suchmaschine ist allerdings zum Teil durch Spezialisierung erkaufte. „Ahoy! The Homepage Finder“ ist für die Suche nach persönlichen Homepages konstruiert (Shakes, Langheinrich & Etzioni, 1998). Als Eingabe verlangt Ahoy! den Namen der gesuchten Person, möglichst auch die Institution, bei der sie beschäftigt ist. Unter Verwendung des Outputs verschiedener Suchmaschinen, E-Mail Verzeichnisse und eigener Datenbanken mit Informationen über Institutionen und URL-Mustern versucht Ahoy! daraufhin die Homepage der entsprechenden Person zu ermitteln. In der Regel werden die ein oder zwei wahrscheinlichsten Treffer aus dem

Output der Suchmaschinen heraus gefiltert, gelegentlich konstruiert Ahoy! aber auch auf der Basis des gesammelten Wissens neue mögliche Adressen, die dann überprüft werden. In entsprechenden Tests erwies sich Ahoy! als recht erfolgreich bei der Identifizierung von Homepages und dürfte besonders zur Suche nach den persönlichen WWW-Seiten von Wissenschaftlern an amerikanischen Universitäten gute Dienste leisten. Eine Reihe weiterer problemspezifischer Softbots zur Informationssuche im Internet werden zur Zeit entwickelt (Softbot Project, 1997).

5.5 Suchmaschinen für E-Mail Adressen

Wie erwähnt, greift Ahoy! bei der Suche nach Homepages unter anderem auf Verzeichnisse mit E-Mail Adressen zurück. Derartige Verzeichnisse sind ein Beispiel für jene Suchmaschinen, die sich darauf spezialisiert haben, E-Mail Adressen zu finden. Das Problem, auf das solche „Personensuchmaschinen“ reagieren, ist das gleiche wie bei WWW-Dokumenten auch: Es gibt keine zentrale Verwaltung sämtlicher E-Mail Adressen und daher auch kein offizielles E-Mail Adreßbuch. Jeder, der einen der 30 Millionen Server (Network Wizards, 1998; RIPE, 1998) betreibt, kann innerhalb weniger Sekunden einen E-Mail Account einrichten. Und genau das geschieht auch. Adressen entstehen in Sekunden und werden oft schon nach kurzer Zeit wieder gelöscht (Karzaunikat, 1998).. Um das Chaos zu komplettieren, gibt es neben den „echten“ auch noch Millionen virtueller E-Mail Adressen. Eine virtuelle E-Mail Adresse leitet Mitteilungen an die „echte“ E-Mail Adresse lediglich weiter. Virtuelle E-Mail Adressen sind sehr beliebt, da sie es den Nutzern zum einen gestatten, ihre Identität zu verbergen. Namen virtueller E-Mail Adressen sind frei wählbar und können jederzeit gelöscht oder geändert werden (z.B. My Own Email, 1998). Zum anderen kann die virtuelle E-Mail Adresse unverändert bleiben, auch wenn sich die echte E-Mail Adresse beispielsweise wegen eines Wechsels zu einem günstigeren Provider ändert. Fazit: Wie die Suchmaschinen für WWW-Dokumente, so können auch die Suchmaschinen für E-Mail Adressen niemals den aktuellen Stand der E-Mail Adressen wiedergeben.

Es stellt sich die interessante Frage, wie die Suchmaschinen zu den vielen Millionen E-Mail Adressen kommen. Die Arbeitsweise der Suchmaschinen für E-Mail Adressen ähnelt stark dem Vorgehen der Maschinen, die Seiten des World Wide Web indexieren (vgl. Karzaunikat, 1998). Sie benutzen Robots (Programme, die das Internet durchkämmen), um innerhalb von Webseiten und in Newsgroups nach den charakteristischen Zeichenketten mit dem @ in der Mitte zu fahnden. Eine E-Mail Adresse allein nützt allerdings wenig, wenn nicht weitere Informationen wie der Name des E-Mail Inhabers verfügbar sind. Die Abkürzungen der E-Mail Adressen selber lassen nur selten Rückschlüsse auf den Inhaber zu. Die Software muß also das Umfeld einer E-Mail Adresse auswerten können, um neben der Zeichenkette der Adresse die Zusatzinformationen herauszufiltern..

Jeder, der Webseiten im Internet mit seiner E-Mail Adresse veröffentlicht oder an Diskussionen in Newsgroups teilnimmt, kann theoretisch in den Datenbanken der Suchmaschinen landen ohne sich je bei der Maschine angemeldet zu haben. Nicht verschwiegen werden sollte, daß mit der Erfassung in Datenbanken - ob nun automatisch oder manuell per Selbsteintrag - auch Nachteile verbunden sein können. Solche Daten können und werden mißbraucht. Insbesondere können mit den Daten vollautomatisch Werbebotschaften an ein Millionenpublikum verschickt werden. Leider ist dies keine Ausnahme, sondern vielmehr Alltagserfahrung. Auch wenn viele seriöse Betreiber von Suchmaschinen glaubwürdig beteuern, ihre Datenbestände nicht an Interessenten aus der Werbewirtschaft weiterzuverkaufen, ist man vor den automatischen Robotern nicht sicher, die im Prinzip jeder mit dem Ziel, dem Internet seine Millionen E-Mail Adressen zu entlocken, installieren kann. Mittlerweile gehen immer mehr Internetnutzer dazu über ihre E-Mail Adresse zu maskieren, damit die automatischen Roboter keine Chance mehr haben. Der Fantasie bei der Art der Maskierung sind dabei keine Grenzen gesetzt. Beispielsweise gibt der Autor einer Webseite seine E-Mail Adresse als ?hahn??@?psycho?logie.de?? an, mit einem Verweis auf eine Fußnote, in welcher erläutert wird, daß bei Nutzung der E-Mail Adresse alle Fragezeichen wegzulassen sind. Die Suchroboter können diese Maskierungen natürlich nicht auswerten und der Autor ist einigermaßen sicher vor den ungeliebten Werbemails. Von den weltweit geschätzten 107 Millionen Internetnutzern (Nua, 1998) treten die meisten nur als Sender und Empfänger von E-Mails in Erscheinung. Da die E-Mail Kommunikation nur nach Einrichtung spezieller Software vor Ort abhörbar ist, können die hier ausgetauschten E-Mail Adressen nicht durch Roboter automatisch erfaßt werden.

Welche Möglichkeiten bestehen nun, auch ohne Rückgriff auf ein zentrales Benutzerverzeichnis die E-Mail Adresse eines bestimmten Teilnehmers der 107 Millionen Internetnutzer in Erfahrung zu bringen? Das Internet bietet eine Vielzahl verschiedener Suchhilfen: *E-Mail Verzeichnisse* unterschiedlichster Größe im World Wide Web, *Usenet-Adressen-Dienste*, *Unternehmens- und Institutsverzeichnisse*, das weltweite *X.500 Verzeichnis*, *Whois-Dienste*, den Service *Netfind* und das Suchwerkzeug *Finger* (vgl. Koopmann, 1997). Diese, in der Reihenfolge ihrer Bedeutung für die Recherche genannten Verzeichnisse und Dienste werden zusammenfassend als *White Pages* bezeichnet. Die Vielzahl der Dienste garantiert jedoch keinesfalls den Erfolg der Suche.

Die größten *Datenbanken mit E-Mail Adressen* sind *im World Wide Web* zu finden. In der Regel sollte die Nutzung der hier zur Verfügung gestellten Dienste ausreichen. Dienste wie das X.500 Verzeichnis oder Whois erfordern Informationen, die über Vor- und Zunamen, sowie wahrscheinlichem Wohnort der gesuchten Person, hinausgehen. Im Web gibt es zum einen E-Mail Verzeichnisse, die auf den Eintragungen der Netznutzer beruhen und zum anderen Verzeichnisse, die zusätzlich Robots eingesetzt haben. Erstere Art verfügt über relativ wenige Adressen (bis zu 100.000, z.B. suchen.de, mail.at, SwissInfo) und verzeichnet nur regionale Adressen. Datenbanken der zweiten Art verzeichnen 10 Millionen und mehr internationale Adressen. Viele der

größeren Datenbanken finanzieren den Betrieb über ein gehöriges Maß an Online-Werbung. Die wichtigsten Datenbanken sind in Tabelle 5.4 zusammengestellt und kurz kommentiert.

Tabelle 5.4: E-Mail Adressendatenbanken im Web

Deutschsprachige Verzeichnisse

Bigfoot Deutschland (<http://www.bigfoot.de/>)

Deutscher Ableger von bigfoot.com. „Das größte Verzeichnis mit den meisten E-Mail-Adressen im Internet“. Datenbestand: 8 Millionen

mail.at - White Pages (<http://www.mail.at/>)

Österreichisches Verzeichnis mit über 8.000 Einträgen

NCC's E-Mail und Homepage Register (<http://www.osthessen.com/emaildb/>)

E-Mail Register, ermöglicht Personen, Firmen, Institutionen in Deutschland und dem Rest der Welt zu finden. Datenbestand: unbekannt

Suchen.de (<http://www.suchen.de/>)

100.000 Einträge einschließlich Verweise auf persönliche Homepages (Telekom-Tochterunternehmen)

SwissInfo (<http://www.swissinfo.ch/email/>)

Schweizer E-Mail-Verzeichnis. Datenbestand: unbekannt

World Wide Internet Phonebook (<http://www.super.de/phone/>)

Keine Informationen zur Datenbank verfügbar

Internationale Verzeichnisse

ESP E-Mail Search Program (<http://www.esp.co.uk/>)

Europäische Suchmaschine erlaubt einfache Suchanfragen: „Bill G“ listet alle Bills auf, deren Nachname mit G beginnt. Datenbestand: 2,5 Millionen

Four11 (<http://www.four11.com/>)

Neben E-Mail Adressen sind persönliche Homepages, Internet-Telefonie- und Videotelefon-Listen verzeichnet. Special: PGP-Schlüssel zum Versand verschlüsselter E-Mails. Gehört mittlerweile zu *Yahoo!* und ist mit *Yahoo! People Search* identisch. Datenbestand: über 10 Millionen

Infospace (<http://www.infospace.com/>)

Server mit 25 großen Datenbanken. E-Mail Suche ist auf bestimmte Domains einschränkbar. Bietet wie IAF auch eine umgekehrte Suche. Besonderheit: die Datenbanken sind untereinander vernetzt. Datenbestand: keine Angabe

Internet Address Finder / IAF (<http://www.iaf.net/>)

Sehr schnelles Suchsystem. Umgekehrte Suche möglich: E-Mail Adresse liefert Information zu der Person. Trunkierungen werden unterstützt (z.B. Hah*). Datenbestand: 6,7 Millionen

OKRA net.citizen Directory Service (<http://okra.ucr.edu/okra/>)

Suchfunktionen (Fuzzylogik): Vor- und Nachnamen, Phrasen, Namen von Organisationen, Adressen von Internet-Computern, Orte, E-Mail Adressen in beliebigen Kombinationen verwendbar. Pro Suchanfrage werden maximal 50 Treffer ausgegeben. Datenbestand: 5,5 Millionen

Switchboard (<http://www.switchboard.com/>)

Suchmaschine für E-Mail Adressen, Websites, Telefonnummern, Firmen. Einfache und schnelle Suche. Datenbestand: keine Angabe

WhoWhere? (<http://www.whowhere.com/>)

E-Mail Adressen von Privatpersonen und Firmen, sowie viele weitere Datenbanken. Die implementierte „Fuzzy-Logic“ führt schon dann zu Treffern, wenn Übereinstimmungen von zwei Buchstaben innerhalb des Namens vorhanden sind. Datenbestand: 16 Millionen

World Email Directory / WED (<http://www.worldemail.com/>)

Zugriff auf Millionen von E-Mail- und Unternehmensadressen sowie Telefonnummern. Suchfunktionen: Eingabe von Namen (Personen, Unternehmen, Organisationen), Teile einer E-Mail Adresse, von Postleitzahlen oder von Telefonnummern. Nicht registrierten Benutzern werden nie mehr als 10 Treffer rückgemeldet. Datenbestand: 18 Millionen

Meta-Suchmaschinen

MESA, MetaEmailSearchAgent (<http://mesa.rrzn.uni-hannover.de/>)

Recherchiert gleichzeitig in den Datenbanken: Populus, IAF, Infospace, suchen.de, SwisInfo, Four11 und Bigfoot

PeopleSearch (<http://www.w3com.com/psearch/>)

Recherchiert in bis zu vier von sechs auswählbaren E-Mail Datenbanken gleichzeitig: Four11, Bigfoot, IAF, WhoWhere, MIT, Whitepages

Die in Tabelle 5.4 aufgeführten Datenbanken sind alle über einfache Formulare abzufragen. Besonders empfehlenswert ist die seit Mai 1997 verfügbare Meta-Suchmaschine **MESA (MetaEmailSearchAgent)** des Regionalen Rechenzentrums für Niedersachsen. Die Suchmaschine befragt gleichzeitig bis zu sieben E-Mail-Datenbanken. Da hier sowohl die kleineren nationalen als auch die großen internationalen Datenbanken kombiniert werden können, ist MESA sicherlich zur Zeit die beste Einzelempfehlung. Bei der Suche sind keine Angaben außer Name und Ort möglich, so daß dieser Service auch ganz besonders für Einsteiger geeignet ist.

Neben diesen großen komfortablen Datenbanken stehen weitere fünf bis sechs Dienste zur Verfügung, mit deren Hilfe sich E-Mail Adressen ermitteln lassen (vgl. Tabelle 5.5). Eine Fundgrube ist beispielsweise das *Usenet* mit seinen mehr als 20 000 Diskussionsgruppen. Wenn davon ausgegangen werden kann, daß die gesuchte Person sich an einer der Gruppen mit eigenen Beiträgen beteiligt hat, dann ist die Wahrscheinlichkeit sehr groß, die E-Mail Adresse mit einer der vier in Tabelle 5.5 aufgeführten Maschinen auch zu finden. Wegen des größten und am längsten zurückreichenden Datenbestandes ist hier insbesondere **DejaNews** zu empfehlen.

Tabelle 5.5: Weitere Informationsdienste zur Suche nach E-Mail Adressen

Adressen aus dem Usenet

AltaVista (<http://www.altavista.digital.com/>)

In der Usenetsuche kann über den Befehl „from:Vorname Nachname“ in allen Beiträgen der letzten 6 Monate recht erfolgreich recherchiert werden.

DejaNews (<http://www.dejanews.com/>)

Der webbasierte Usenet-Server ermöglicht mit der Option „Powersearch“ ebenfalls die Recherche nach den Autoren in 100 Millionen Postings seit dem Jahr 1995.

Netaddress Finder (<http://www.obh.snafu.de/cgi-bin/netaddress>)

Der Service von Oliver Brandmüller ist speziell für die Suche im deutschsprachigen Usenet geeignet. Alle Absender der letzten 42 Tage.

Usenet Addresses Database (<http://usenet-addresses.mit.edu/>)

Gesucht wird im Bestand von 4,1 Millionen E-Mail Adressen, die zwischen Juli 1991 und Februar 1996 im Usenet gepostet wurden.

Adressen mit WhoIs

InterNIC WhoIs (<http://rs.internic.net/cgi-bin/whois>)

Daten zu allen Internetcomputern und ihren Betreibern der Top-Level-Domains .com, net und .org

RIPE WhoIs (<http://www.ripe.net/db/whois.html>)

Daten zu allen Internetcomputern und Ihren Betreibern europäischer Top-Level-Domains.

Adressen mit X.500

X-500 Auffahrt des DFN-Vereins (<http://ambixhp2.uni-tuebingen.de:10111/>)

X-500 Auffahrt der Universität Gießen (<http://www-x500.uni-giessen.de:8890/>)

X-500 Auffahrt der Universität Chemnitz (<http://tricia.hrz.tu-chemnitz.de:8888/>)

Die X-500 Auffahrt der Universität Chemnitz verfügt über das beste Webinterface.

Adressen mit Netfind (login: netfind)

Netfind-Server der José State University, California, USA (**telnet:**netfind.sjsu.edu)

Netfind-Server des Imperial College, London, England (**telnet:**monolith.cc.ic.ac.uk)

Netfind-Server des InterNIC Directory and DB Services, S. Plainfield, NJ, USA (**telnet:**ds.internic.net)

Viele Institutionen und Unternehmen unterhalten sogenannte *Whois Server*, auf denen die Adressen ihrer Mitarbeiter oder Nutzer gespeichert sind. Whois Server gibt es auf der ganzen Welt. Allerdings nimmt die Zahl laufend ab und in Deutschland hat es ohnehin nie viele gegeben. Die meisten deutschen Whois Server werden von Universitäten betrieben. Whois Server lassen sich leider nur in seltenen Fällen über das World Wide Web befragen. In der Regel können sie nur über Telnet oder mit speziellen Whois Clienten abgefragt werden. Die offizielle Whois Datenbank des InterNIC ist auch im Web erreichbar. In dieser Datenbank sind die Domainnamen für die Top-Level-Domains .com, .net und .org gespeichert, nebst Informationen zu den Betreibern der Internetserver wie Namen, Anschriften, Telefon-, Fax- und E-Mail Nummern sowie den IP-Adressen der Server. Sind also Informationen über den Arbeitgeber einer gesuchten Person verfügbar, so läßt sich mit dieser Datenbank zumindest der Verwalter des Rechners des Unternehmens oder der Institution per E-Mail erreichen. Dieser kann in der Regel feststellen, ob die gesuchte Person über eine E-Mail Adresse verfügt. Eine Liste fast aller im Internet zu findenden Whois Server findet sich bei Mike Power (1998).

Ähnlich wie im Falle der Whois Server kommt man auch über den *X.500 Dienst* nur selten direkt zum Ziel. X.500 ist der Versuch, ein international standardisiertes E-Mail

Verzeichnis aufzubauen. Die X.500 Server, von denen es ein paar hundert weltweit gibt, sind untereinander vernetzt. Bei Nutzung etwa des in Tabelle 5.5 angegebenen Servers des DFN-Vereins steht also die gesamte Datenbank zur Verfügung. Die Suche erfolgt mit Hilfe einer Hierarchie, die nur auf der ersten Ebene geographisch verzweigt (Welt->Staaten). Auf der Ebene eines Staates (z.B. Germany) werden dann bereits alle Institutionen oder Unternehmen angezeigt, die ein X.500 Verzeichnis unterhalten (169 in Deutschland). Von diesen Organisationen werden (oder vielmehr sollten) weitere Unterorganisationen sowie die Namen und Adressen der Mitarbeiter gespeichert werden. An der Humboldt-Universität zu Berlin sind beispielsweise lediglich der Fachbereich Informatik, das Rechenzentrum und der Postmaster bekannt. Nicht sehr viel für eine Institution an der ein paar Tausend Menschen beschäftigt sind. Ähnlich wie im Falle der Whois Server bleibt keine andere Möglichkeit als den Postmaster per E-Mail um die gewünschte Information zu bitten. Da viele Universitäten auf ihren Homepages im Web Mitarbeiterverzeichnisse anbieten, empfiehlt sich in der Regel deren Nutzung.

Als letzte Möglichkeit zur Ermittlung einer E-Mail Adresse sei das fast vergessene Suchwerkzeug *Netfind* genannt. Für die Suche mit Netfind wird nicht nur der Name der gesuchten Person benötigt. Zusätzlich muß der E-Mail-Server ermittelt werden, auf dem die gesuchte Person wahrscheinlich ein Postfach hat. Um eine Liste möglicher E-Mail-Server zu erstellen, erlaubt Netfind auch Suchangaben wie Städte-, Orts-, Länder- oder Organisationsnamen. Netfind versucht also zunächst diejenigen E-Mail-Server zu finden, in der der Gesuchte ein Postfach haben könnte. Hierbei greift Netfind auf die Whois Datenbank von InterNic sowie auf X.500 Server zurück. Die eigentliche Anfrage nach der gesuchten Person stellt Netfind dann aber mittels Finger und SMTP an die ausgewählte Liste von Ziel-E-Mail-Servern. Da Netfind keine eigene Datenbank unterhalten muß, sondern einfach die E-Mail-Server fragt, ob eine bestimmte Person aktuell über ein Postfach verfügt, wäre Netfind eigentlich ein hervorragendes Tool. Leider setzt Netfind voraus, daß auf den zu befragenden E-Mail-Servern der Dienst Finger läuft. Finger wurde aber auf sehr vielen E-Mail-Servern aus Datenschutzgründen inaktiviert. Mit dem Dienst *Finger*, der auch als eigenständiges Clientprogramm für viele Computertypen und über das World Wide Web verfügbar ist, kann jeder Internetnutzer einen bekannten E-Mail-Server (z.B. mail.fu-berlin.de) fragen, ob eine bestimmte Person ein Postfach hat (z.B. „finger Nachname_der_gesuchten_Person@mail.fu-berlin.de“). Ist Finger auf dem E-Mail-Server abgeschaltet, dann bleiben die Anfragen von Netfind unbeantwortet und die Suche erfolglos. Netfind hat sich wahrscheinlich bis heute auch deshalb nicht durchgesetzt, weil die Netfind Server ausschließlich über Telnet (vgl. Schade, in diesem Band) zu erreichen sind (Login: netfind; Passwort ist keines erforderlich). Bis Dezember 1997 stand auch ein WWW-Gateway zur Verfügung, daß der Autor allerdings wegen zahlreicher technischer Probleme wieder geschlossen hat. Wer keine Probleme mit der Nutzung eines Telnetclients hat, sollte einen der in Tabelle 5.5 aufgeführten Server

testen. Ein Beispiel für eine Netfindsitzung geben Kessler und Shepard (1994, RFC Dokument Nummer 1739, Abschnitt 2.7.3).

5.6 Suchmaschinen für FTP-Server und Softwaredatenbanken

Dokumente beliebiger Formate, Graphiken, Animationen, Videofilme, Musikstücke und vor allem Computerprogramme werden nicht nur im World Wide Web zur Verfügung gestellt, sondern auch auf tausenden von öffentlich zugänglichen, sogenannten „anonymen“ FTP-Servern rund um den Globus. Ein FTP-Server ist ein an das Internet angeschlossener Computer, der die Möglichkeit bietet, mit Hilfe einer speziellen Software Dateien von einem Besucher zu empfangen (upload) oder umgekehrt, Dateien an den Besucher zu senden (download). Wie schon der ausführliche Namen verrät, geht es beim **File Transfer Protocol** also primär um das Fernkopieren von Dateien (Programmen, Daten) und nicht etwa, wie beim Surfen durchs WWW, um die Darstellung des Dateiinhalts (vgl. Schade, in diesem Band). Der Blick in einen FTP-Server gleicht dann auch dem Blick in die Dateihierarchie auf der lokalen Festplatte (vgl. Karzaunikat, 1998). Was der Besucher sieht, sind Verzeichnis- und Dateinamen, sowie Angaben zu den Dateien (Datum, Größe, Attribute). Um ein Dokument oder ein Programm von einem FTP-Server übertragen zu können, muß der Name des FTP-Servers, der Verzeichnispfad und der Name der Datei bekannt sein (z.B.: ftp.fu-berlin.de/pub/windows98/winzip.zip). Was aber tun, wenn zwar der Dateiname eines Dokuments oder Programms ungefähr bekannt ist, aber keine Information verfügbar ist, ob es einen anonymen FTP-Server gibt, der genau diese Datei bereithält?

Um Dateien auf FTP-Servern zu lokalisieren wurden drei Suchmaschinen entwickelt, die die Inhaltsverzeichnisse der weltweit frei zugänglichen FTP-Server verzeichnen: **Archie**, **FTP Search** und **Filez** (vgl. Tabelle 5.6). Archie und FTP Search verzeichnen lediglich Datei- und Verzeichnisnamen. Inhaltsangaben zu den Dokumenten (so weit vorhanden) sind nur bei Filez recherchierbar. Bei der Suche nach einer Software mit unbekanntem Datei- oder Archivnamen braucht man also etwas Glück, um das Verzeichnis, in dem die gesuchte Datei liegt, zu finden. Bei Computerprogrammen ist die Suche noch vergleichsweise einfach, da die Dateinamen zumindest teilweise dem Namen des Programms entsprechen. Textdokumente sind dagegen schwierig oder gar nicht zu finden.

In Archie oder FTP Search kann der Benutzer den Dateinamen des gesuchten Programms oder ein kleines Fragment des Dateinamens eingeben und erhält als Rechercheergebnis eine Liste mit FTP-Servern, von denen die Datei geladen werden kann. Die Rückmeldung umfaßt dabei auch Angaben zum Verzeichnis, in dem die Datei auf dem Server abgelegt wurde und Informationen über Datum und Größe der Datei. Minimale Übertragungszeiten erreicht man durch Wahl eines geographisch möglichst nahen Servers. Der Zugang zu Archie, Filez und FTP Search ist komfortabel über das World Wide Web möglich. Die Technische Hochschule Darmstadt betreibt den deut-

schen Archie-Server. Dort finden sich auch detaillierte Anleitungen, Suchtips und ein Verzeichnis öffentlicher deutscher FTP-Server. Für den Suchdienst Archie gibt es auch spezielle Programme, die für nahezu alle Computertypen und Betriebssysteme zur Verfügung stehen. Der norwegische Suchdienst FTP Search ist nur über das WWW zu erreichen, bietet allerdings ausgefeiltere Suchmöglichkeiten. So kann hier beispielsweise auch der Datumsbereich der gesuchten Datei definiert werden. Filez ist die beste Empfehlung, wenn der Name der gesuchten Datei nicht bekannt ist oder nicht erschlossen werden kann und daher auf die Kurzbeschreibungen zurückgegriffen werden muß. Die Suche in Filez nimmt in der Regel mehr Zeit in Anspruch.

Tabelle 5.6 Suchmaschinen und Datenbanken für Software und Dateien

FTP-Suchmaschinen

Archie (<http://archie.th-darmstadt.de/archie.de.html>)

Archie-Server der Universität Darmstadt

(Fast) FTP Search (<http://ftpsearch.ntnu.no/>)

Verzeichnet Dateien von mehr als 7.400 FTP-Servern. Sehr schnell und sehr ausgefeilte Suchmöglichkeiten

Filez (<http://www.filez.com/>)

Verzeichnet Dateien von mehr als 7000 FTP-Servern. Suchfunktionen: Betriebssystem, Dateart (z.B. Video, Audio) und Beschreibungen. Bonus: eigenes Softwarearchiv

Software-Datenbanken

Jumbo (<http://www.jumbo.com/>)

Über 93.000 Free- und Sharewareprogramme

Software.com (<http://www.shareware.com/>)

Über 250.000 Sharewareprogramme getrennt für die diversen Betriebssysteme

Tucows (www.tucows.com/)

Software fürs Internet: sehr gutes Kategoriensystem und Programmbeschreibungen. Jedes Programm direkt von einem der über 300 Mirror Server übertragbar.

32-bit.com (www.32bit.com/)

Software für 32-Bit Betriebssysteme: Zusammenfassung diverser bekannter Archive

LEO - Link Everything Online (www.leo.org/archiv/software/software_d.html)

Umfangreicher deutscher Softwarekatalog

FTP Index Files (www.tosti.com/fsuche/index_de.htm)

Deutscher Suchdienst von Dirk Tostmann: gesucht wird in den Indexen der sechs größten Softwarearchive gleichzeitig.

Neben den FTP-Suchmaschinen finden sich einige sehr umfangreiche *Softwaredatenbanken* für alle gängigen Betriebssysteme im World Wide Web. Diese Datenbanken verzeichnen neben Daten zur Fundstelle im Internet auch Beschreibungen zum Anwendungsbereich der jeweiligen Software. Systeme wie **software.com** von c|net bieten mehr als 250 000 Programme zum direkten Download an. Die wichtigsten und größten Archive können der Tabelle 5.6 entnommen werden. Alle genannten Quellen bieten selbstverständlich nur Software an, die als Freeware, Public Domain, Testware oder Shareware vertrieben wird. Für den Erwerb kommerzieller Softwarepakete ste-

hen zahlreiche Online Softwareshops zur Verfügung (vgl. Tabelle 5.1 für entsprechende Verweislisten). Natürlich sind die meisten Hersteller auch mit eigenen Webservern vertreten (z.B. SPSS, <http://www.spss.com/>). Auf Datenbanken wie SI-Byl oder CTI, die speziell psychologische Software verzeichnen, gehen wir im nächsten Kapitel ein (Hahn & Günther, in diesem Band).

5.7 Weiterführende Informationen

Suchmaschinen sind ein unendliches Thema mit vielen Aspekten, von denen hier nur wenige angesprochen werden konnten. Wer an weiterführenden und vertiefenden Informationen interessiert ist, der kann diese nicht zuletzt im Internet selbst finden. An erster Stelle ist hier die „Die kleine Suchfibel“ von Stefan Karzauninkat (<http://www.suchfibel.de/>) zu nennen, eine vorzügliche, leicht verständliche, an der Suchpraxis orientierte und dennoch gründliche Einführung in den Umgang mit Suchmaschinen. Dieses Online-Buch, das noch 1998 auch in einer Printausgabe erschienen wird und dennoch nicht aus dem Netz verschwinden soll, führt Schritt für Schritt in den Umgang mit Suchmaschinen ein, stellt die wichtigsten Systeme kompetent und sachlich vor und hält darüber hinaus auch für den, der sich für Hintergrundinformationen interessiert, eine Fülle an vertiefenden Texten bereit. Von Suchstrategien bis zur Syntax einer Abfrage erfährt der Leser alles, was er für den Umgang mit Suchmaschinen wissen muß. Doch nicht nur der Informationssuchende, sondern auch derjenige, der Informationen anbietet und deshalb von möglichst vielen Netznutzern gefunden werden möchte, erhält wertvolle Hinweise. Der Schwerpunkt der „kleinen Suchfibel“ liegt auf der Suche nach WWW-Dokumenten, doch werden auch spezielle Suchhilfen behandelt wie Medienarchive, „Menschensuchmaschinen“ oder Nachrichtendienste.

Eine amerikanische, ebenfalls ausschließlich dem Thema Suchmaschinen gewidmete Website ist „Search Enging Watch“ von Danny Sullivan (<http://www.searchenginewatch.com/>). Auch hier gibt es eine Vielzahl an Informationen und jede Menge Links zu anderen Informationen rund um das Thema Suchmaschinen, beschränkt allerdings im wesentlichen auf Volltextsuchmaschinen. Erwähnt sei schließlich noch eine informative Einführung, die im „PC Magazine Online“ unter dem Titel „Your Complete Guide to Searching the Net“ (http://www.zdnet.com/pcmag/features/websearch/_open.htm) erschienen ist und neben Volltext- sowie verzeichnisbasierten Suchmaschinen auch Personensuchmaschinen (White Pages), Firmensuchmaschinen (Yellow Pages), Usenet-Suchmaschinen sowie Meta-Suchmaschinen vorstellt.

5.8 Resümee oder: Wie viele Suchmaschinen braucht der Mensch

Suchmaschinen helfen zweifellos, die Informationsfülle oder, vielleicht besser, das Datenchaos des Internets zu erschließen. Ohne sie, dies gilt es noch einmal zu betonen, wäre das Internet nur die Hälfte wert. Dennoch tragen auch Suchmaschinen ihren Teil zur Verwirrung bei. Verzeichnisse, die sich rühmen, mehr als 3.000 „Suchmaschinen“ zu kennen, Suchmaschinen-Suchmaschinen und Meta-Suchmaschinen, die teilweise ihrerseits wieder Meta-Suchmaschinen abfragen, zeugen hiervon. Der Nutzen einer derartigen Vielzahl von Suchmaschinen und Suchhilfen ist gelegentlich zweifelhaft und droht hinter den Orientierungsverlusten zurückzutreten, die durch eine Pseudo-Differenzierung weitgehend austauschbarer Systeme bewirkt wird. Man muß sich hier vor Augen halten, daß zumindest hinter den großen Suchmaschinen im Netz häufig miteinander konkurrierende Unternehmen stehen, die ein kommerzielles Produkt vermarkten und vor allem zur Erzielung von Werbeeinnahmen an hohen Besucherzahlen interessiert sind. So nahm beispielsweise das Unternehmen Yahoo! Inc. im ersten Quartal 1998 über 30 Millionen Dollar ein (Yahoo!, 1998b), die Infoseek Corporation im selben Zeitraum über 14 Millionen Dollar (Infoseek, 1998). Dieser Aspekt fällt dem Nutzer weniger auf, da er in der Regel für die Inanspruchnahme der Informations- und Kommunikationsdienste dieser Unternehmen nichts zahlen muß. Er erklärt aber, warum häufig bei den Betreibern von Suchmaschinen wenig Interesse an Transparenz und Vergleichbarkeit oder gar Kooperation besteht. Daher sei hier die Frage an den Schluß gestellt, wie viele Suchmaschinen der Netzbenuer überhaupt braucht.

Natürlich gibt es keine allgemeine Antwort auf diese Frage. Die Bedürfnisse der Nutzer sind ebenso verschieden wie ihr Anspruchsniveau bei der Informationsrecherche. Dennoch sollte nicht der Eindruck entstehen, man müsse alle hier vorgestellten Systeme kennen, um mit Aussicht auf Erfolg im Netz recherchieren zu können. Vieles ist austauschbar, manche Maschinen eindeutig besser als andere, was allerdings nicht immer leicht zu erkennen ist. Das grundlegende Konstruktionsprinzip des Internets, das viele Wege zum Ziel führen (vgl. Musch, in diesem Band), gilt im übertragenen Sinne auch für die Informationssuche: Verschiedene Suchmaschinen können zum selben Ergebnis führen und innerhalb desselben Systems kann man auf ganz unterschiedliche Weise relevante Informationen einkreisen. Häufig dürfte es sinnvoller sein, sich mit der Bedienung weniger Suchsysteme vertraut zu machen, als viele von ihnen oberflächlich zu kennen. Fantasie, Kreativität und ein Schuß detektivischer Spürsinn helfen möglicherweise weiter als die Wiederholung immer wieder derselben Abfrage auf weitgehend redundanten Suchmaschinen. Demgegenüber ermöglicht die Kombination *verschiedener* Suchsysteme oft ungewöhnliche Lösungen alltäglicher wissenschaftlicher Aufgabenstellungen.

Um ein kurzes Beispiel zu geben: Nehmen wir an, wir würden im Rahmen einer geplanten Studie eine psychometrische Skala zur Erfassung generalisierter Selbstwirksamkeit suchen. Eine Suchmaschinen liefern zwar auf das Schlagwort „self-efficacy“

zahlreiche WWW Seiten zurück, leider enthält aber keine der durchgesehenen Seiten die gewünschte Information. Auf einer Seite stoßen wir allerdings auf einen Link zu einem „Test Locator“ an der Catholic University of America (vgl. Hahn & Günther, in diesem Band). In der Tat treffen wir dort auf eine umfangreiche Testdatenbank und eine Abfrage fördert über 20 verschiedene Verfahren zu Tage. Nur drei Verfahren scheinen die Möglichkeit zur Erfassung allgemeiner Selbstwirksamkeit zu bieten. Einer der genannten Autoren ist Ralf Schwarzer, dessen Name uns ohnehin schon in diesem Zusammenhang begegnet ist. Wir beschließen, nach seiner E-Mail Adresse zu suchen. Auch diese Suche verläuft nach Einengung auf den „richtigen“ Ralf Schwarzer erfolgreich. Schließlich bitten wir den Autor in einer E-Mail um die Zusendung von Unterlagen. Viele weitere Szenarios sind denkbar.

Eine solche Suche mag zwar umständlich und mühsam erscheinen, dürfte aber für den Alltag der Informationsrecherche im Internet repräsentativer sein als der Glücksfall, bei dem schon die erste Anfrage bei einer Suchmaschine zum Erfolg führt. Ein gewisser Trost liegt vielleicht darin, daß man auf diese Weise nicht selten wertvolle Entdeckungen macht, nach denen man eigentlich gar nicht gesucht hat.

Literatur

- AltaVista (1998a). *About AltaVista*. [WWW document.] URL
http://www.altavista.digital.com/av/content/about_our_story_2.htm
- AltaVista (1998b). *Digital's AltaVista Search index grows to record heights*. Pressemitteilung vom 27. Mai 1998. [WWW document.] URL
<http://www.altavista.digital.com/av/content/pr052798.htm>
- Bharat, K. & Broder, A. (1998a). A technique for measuring the relative size and overlap of public Web search engines. Paper presented at the 7th International World Wide Web Conference (WWW7), April 1998. [WWW document.] URL
<http://www7.conf.au/programme/fullpapers/1937/com1937.htm>
- Bharat, K. & Broder, A. (1998b). *Measuring the Web*. [WWW document.] URL
<http://www.research.digital.com/SRC/whatsnew/sem.html>
- Infoseek (1998a) *Infoseek introduces „E.S.P.“ to dramatically improve general search results*. Pressemitteilung vom 20. Mai 1998. [WWW document.] URL
<http://info.infoseek.com/doc/PressReleases/esp.html>
- Infoseek (1998b). *Infoseek reports 131% growth in first quarter revenue*. Pressemitteilung vom 23. April 1998. [WWW document.] URL
http://info.infoseek.com/doc/PressReleases/131_percent.html
- Kalb, D. (1998). E-Mail von Detelf Kalb, Projektleiter Fireball/Paperball, vom 5. Mai 1998
- Karzauninkat, St. (1998). *Die kleine Suchfibel. Wie findet man Informationen im Netz?* [WWW document.] URL <http://www.suchfibel.de/>
- Kessler, G. & Shepard, S. (1994, December). A Primer On Internet and TCP/IP Tools [Request for Comments: 1739]. Hill Associates, Inc. [WWW document.] URL
<http://www.internic.net/help/domain/rfc1739.txt>

- Koopmann, J. (1997). *E-Mail Adressen Suchen 4.1.1*. [WWW document.] URL http://ourworld.compuserve.com/homepages/Jonas_/email/suche.html
- Lawrence, St. & Giles, C. L. (1998). Searching the World Wide Web. *Science*, 280, 98-100
- Media Metrix (1998). *Media Metrix. The PC Meter Company*. [WWW document.] URL <http://www.mediametrix.com/>
- Monz, G. (1998). E-Mail von Gerhard Monz, Intersearch Europe, vom 3 Juni 1998
- My Own Email (1998). *200 personalized domain names!* [WWW document.] URL <http://www.myownemail.com/>
- Network Wizards (1998, April). *Internet Domain Survey*. Menlo Park, California: Network Wizards. [WWW document.] URL: <http://www.nw.com/zone/WWW/top.html>
- Nua (1998, April). *How many online?* Ireland: Nua Ltd. [WWW document.] URL http://www.nua.ie/surveys/how_many_online/
- Power, M. (1998). *List of Internet whois servers, 25 January 1998 (Version 4.9)*. [FTP document.] URL <ftp://sipb.mit.edu/pub/whois/whois-servers.list>
- Rabinovitch, E. (1997) *Metasearch*. PC Magazine online. [WWW document.] URL <http://www.zdnet.com/pcmag/features/websearch/ms.htm>
- RIPE (1998, April). *European Hostcount*. Amsterdam, The Netherlands: Network Coordination Centre. [WWW document.] URL: <http://www.ripe.net/statistics/hostcount.html>
- Schoebel, M. (1998). E-Mail von Michael Schoebel, Suchmaschine Eule, vom 2. Juni 1998
- Shakes, J., Langheinrich, M. & Etzioni, O. (1998). Dynamic reference sifting: A case in the homepage domain. [WWW document.] URL http://www.cs.washington.edu/homes/jshakes/ahoy_paper/paper.html
- Softbot Project (1997). *Softbots. Intelligent software agents for the internet*. University of Washington, Department of Computer Science and Engineering. [WWW document.] URL <http://www.cs.washington.edu:80/research/projects/softbots/www/softbots.html>
- Sullivan, D. (1998a). *Yahoo special report*. [WWW document.] URL <http://www.searchenginewatch.com/webmasters/yahoo/>
- Sullivan, D. (1998b). *Search Engine Alliances Chart*. [WWW document.] URL <http://www.searchenginewatch.com/reports/alliances.html>
- Sullivan, D. (1998c). *Metacrawlers and Metasearch Engines*. [WWW document.] URL <http://www.searchenginewatch.com/facts/metacrawlers.html>
- Yahoo! (1998a). *Lern about Yahoo!* [WWW document.] URL <http://howto.yahoo.com/chapters/7/1.html>
- Yahoo! (1998b). *Yahoo! Reports Profitable First Quarter Results*. Pressemitteilung vom 8. April 1998. [WWW document.] URL <http://www.yahoo.com/docs/pr/release166.html>