

Für die Praxis

Ferenc Becker

Internet-Suchmaschinen

Funktionsweise und Beurteilung*

Nach einer kurzen Darstellung der Entwicklung der Suchdienste wird das Suchretrieval von AltaVista näher beschrieben. Es werden Kriterien für die Beurteilung der Suchdienste festgelegt. Danach folgt die Beschreibung der Möglichkeiten der einzelnen Suchdienste und ihre Beurteilung. Zum Schluß wird über die Auswahl eines Suchdienstes diskutiert.

Internet Search Engines: function and evaluation

After a short description of the development of search engines the retrieval of AltaVista is described. Criteria for the evaluation of search engines are settled. A description of the possibilities offered by various search services follows together with their evaluation. Finally the selection of a search service is discussed.

Moteurs de recherches sur Internet: fonction et évaluation

Après une description brève du développement des moteurs de recherche on décrit la recherche d'AltaVista. Les critères pour une évaluation des moteurs de recherche sont déterminés. Ensuite on donne une description des possibilités de divers moteurs de recherche et leur évaluation. A la fin on discute la sélection d'un moteur de recherche.

Einleitung

Das Multimediale Internet gewinnt heute immer mehr an Bedeutung. Schlagworte wie *Datenhighway* oder *Informationsflut* hört und liest man überall. Der Internetbenutzer muß sich im Informationsdschungel zurechtfinden. Wegen des dezentralen Aufbaus des Webs ist das Suchen und Finden mit einigen Schwierigkeiten verbunden. Vor einigen Jahren erschienen im Internet Verzeichnisse und Suchmaschinen, um die Suche und die Navigation im WWW zu erleichtern. Inzwischen gibt es viele Suchmaschinen, die zum Teil nationale, internationale und spezielle Bedürfnisse befriedigen. Den Benutzern wird eine Vielzahl an Suchdiensten und die daraus resultierende Vielfalt des Suchretrievals zugemutet.

Um vernünftige Ergebnisse, d.h. eine verhältnismäßig kleine Zahl an möglichst relevanten Treffern zu erreichen, müssen vor allem die Fachleute wissen, an welche Suchmaschine sie sich wenden, wobei sie natürlich auch die Suchmöglichkeiten der Dienste beherrschen müssen. Das Zurechtfinden im Internet bereitet nicht nur Laien sondern auch konventionellen Informationsvermittlern Probleme.

1 Vorüberlegungen zum Thema Suchen

1.1 Hypertext und Hypermedia

Im Zusammenhang mit dem Internet hört man oft vom *Hypertext*. Die Grundidee von Hypertext wäre die Informationen in kleine in sich geschlossene Häppchen zu zerkleinern und diese an den geeigneten Stelle unter-

einander zu verbinden. Diese Verbindungen (*links*) sollen die inhaltlichen Zusammenhänge zwischen den einzelnen Informationshäppchen darstellen. Bei der Verbreitung des Internets entstand auch der Begriff *Hypermedia*, die die multimediale Aufbereitung von Hypertext bezeichnet.

Das Internet selbst sollte ein Hypertext- oder Hypermedia-System sein. Doch wegen der Vielzahl an Autoren im Internet hält sich dieses Medium nicht an diese Grundgedanken. So werden wir im Internet Dokumente finden, die den Grundprinzipien entsprechend erstellt wurden, aber auch sehr viele Dokumente, die ähnlich den gedruckten Medien längere Textpassagen (wie etwa mehrere Kapitel oder gar eine komplette Publikation) enthalten. All dies macht das Suchen in diesem Medium schwierig.

Mit der wachsenden Zahl der Dokumente im Internet wurde gleich in den Anfangsjahren die Erschließung dieses Mediums eine immer dringendere Aufgabe. Zuerst haben die Benutzer das nur für sich realisiert. Das machen wir bis heute, indem wir Verweise auf die interessantesten oder wichtigsten ‚*Orte*‘ im Internet legen. Das sind die sogenannten *Lesezeichen* oder *Bookmarks* bei den Internet-Browsern.

Die Benutzer haben schnell erkannt, daß wegen der großen Menge der Lesezeichen auch deren Systemati-

* Überarbeitete Ausgabe der Diplomarbeit im Studiengang Öffentliche Bibliotheken bei der Fachhochschule Stuttgart, Hochschule für Bibliotheks- und Informationswesen. Ich danke meinen Eltern für die unermüdliche Geduld und Unterstützung.

sierung erforderlich ist. Deswegen entstanden bei den Benutzern die individuellsten Systematiken, die solange diese nur für den persönlichen Gebrauch gedacht waren, gar keine Probleme bedeuteten. Diese persönlichen Systematiken wurden aber bald auch der Öffentlichkeit zugänglich gemacht, und die systematischen Internetkataloge entstanden. Aus ihrer Geschichte heraus, und weil wohl nur wenige oder gar keine Bibliothekare daran beteiligt waren, hat heutzutage jeder systematische Internetkatalog eine eigene Systematik. Diese Systematiken sind meistens auch den im Internet anzutreffenden und gesuchten Informationen angepaßt, und so existieren Systematiken, die von den der Bibliothekskatalogen stark abweichen. In Bibliotheken werden Systematiken benutzt, die den Anspruch haben, das Weltwissen systematisieren zu können, im Internet trifft man Systematiken an, die den Anspruch haben, die Welt des Internets systematisieren zu können.

1.2 Schlagwortrecherche

Das Internet ist ein sehr dynamisches Medium, daher sind systematische Internetkataloge sehr menschen- und arbeitsintensiv. Durch das schnelle Wachstums des Mediums wurde deutlich, daß man die „Katalogisierungsarbeiten“ Rechnern überlassen sollte. So entstanden dann die Schlag- oder Stichwortkataloge. Diese Kataloge gehen selbständig vor und durchsuchen das Internet mithilfe von sogenannten Suchrobotern und nehmen die gefundenen Dokumente (oft in Volltext) in ihren Bestand auf.

In der Umgangssprache werden sowohl systematische als auch Schlagwortkataloge als Suchmaschinen bezeichnet, obwohl nur letztere diese Bezeichnung verdienen. Beide Arten der Suchdienste haben erkannt, daß die Benutzer sowohl systematische als auch Schlagwortkataloge benutzen, so kann man heutzutage sehen, daß die beiden Arten immer mehr ‚aufeinander zugehen‘. Schlagwortkataloge haben sich auch einen systematischen Zugang verschafft und auch umgekehrt. Ich möchte vor allem auf die Benutzung der Schlagwortkataloge näher eingehen, da diese in der Benutzung wesentlich höhere Anforderungen an den Kenntnisstand des Benutzers stellen. Außerdem kann man mit systematischen Katalogen sinngemäß umgehen.

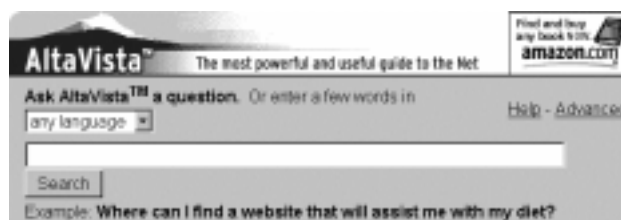
2 Die Schlagwortsuche

In diesem Kapitel werde ich nur auf die Suchmöglichkeiten bei AltaVista eingehen, weil diese die umfangreichsten sind, und weil diese Kenntnisse dann ohne größere Umstellung auf andere Suchdienste übertragbar ist. Bei den meisten Schlagwortkatalogen gibt es einen einfachen (bei AltaVista *Simple Search*) und einen professionellen (bei AltaVista *Advanced Search*) Suchmodus. Fangen wir mit der ‚einfachen‘ Suche an.

2.1 Einfache Suche

Schauen wir uns erst die einfache Suchoberfläche von AltaVista an.

Unter dem Logo von AltaVista ist ein kleines Dropdown-Menü zum Auswählen der Sprache der gesuchten Dokumente angebracht. Daneben kann man zu der Hilfe-



stellung oder zu der Profi-Suche übergehen. Darunter ist das Eingabefeld zum Formulieren der Suchfrage zu finden. Es folgt der Search-Knopf, womit die Suche abgeschickt wird und dann eine freundliche Empfehlung vom Suchdienst, daß man ausformulierte Fragesätze versuchen sollte.

2.1.1 Suche nach einem Schlagwort

z.B.: *linguistik* in German 8458 Treffer
Es ist eindeutig, daß wir zu viele Treffer erhalten hatten.

2.1.2 Suche nach mehreren Schlagworten

z.B.: *linguistik psychologie* in German 64 845 Treffer
Damit haben wir unsere Situation nicht verbessert, eher verschlimmert. Es ist zu sehen, daß AltaVista nicht nach Dokumenten sucht, die *linguistik* und *psychologie* enthalten!

2.1.3 Die Benutzung von + und –

Wenn man (*ohne Lehrzeichen*) vor ein Suchbegriff ein Pluszeichen setzt, erzwingt man, daß dieser Begriff in allen Trefferdokumenten vorkommt. Bei der Benutzung des Minuszeichens darf der Begriff, dem es vorgestellt wurde, in keinem der Trefferdokumente vorkommen.

z.B.: *+linguistik +psychologie* in German 7830 Treffer
Es ist zu sehen, daß die Treffermenge der Suche nach *linguistik* durch den hinzugefügten zweiten Begriff (um 628 Treffer) eingegrenzt wurde.

Oder: *+linguistik +psychologie* in German 3630 Treffer
–philosophie

Es ist sehr wichtig zu verstehen, das die Plus- und Minuszeichen nicht wie in der Mathematik zwei Begriffe miteinander verbinden, sondern sie werden bei Suchmaschinen zur *positiven oder negativen Betonung der Begriffe* verwendet. Will man Dokumente erhalten, die beide Suchbegriffe enthalten sollen, so müssen *beiden* Begriffen Pluszeichen vorgestellt werden!

2.1.4 Phrasensuche

In diesem Fall bedeutet Phrase eine Zeichenkette, die Zeichen für Zeichen genau (auch unter Beachtung der Groß- und Kleinschreibung) gesucht wird. Solche Phrasen werden in Anführungszeichen gesetzt:

z.B.: *„Expo 2000“* in German 1807 Treffer

2.1.5 Groß- und Kleinschreibung

Laut Hilfestellung von AltaVista werden klein geschriebene Worte in allen möglichen Varianten (*altavista, AltaVista, ALTAVISTA, AITaVisTA ...*), und groß, gemischt oder mit Umlauten und Accents geschriebene als Phrasen gesucht.

2.1.6 Trunkierung

Mit einem Stern unmittelbar nach dem Suchbegriff angebracht kann man die Suchworte trunkieren. Das bedeutet, daß alle Worte als Treffer gedeutet werden, die mit dem Stummel vor dem Stern anfangen.

z.B.: *philoso**

Hierbei werden *Philosophie, Philosoph, Philosophen, philosophisch, philosophische, philosophischer, philosophisches, philosophischen, philosophischem*, usw. gefunden.

Oder ein Beispiel für Innentrunkierung:

z.B.: *schlaum*r*

Es werden dabei u.a. folgende Worte gefunden: *Schlaumeier, Schlaumaier, Schlaumacher*.

2.1.7 Suche mit einem ausformulierten Fragesatz

z.B.: *Was ist die Hauptstadt in German 629 775 Treffer der Bundesrepublik Deutschland?*

Eine erschreckende Anzahl an Treffern, die allerdings beim näheren Hinsehen doch als relevant entpuppen. Unter den ersten 10 Treffern sind zwar auch ein paar ‚tote Links‘, aber die restlichen Dokumente enthalten mindestens einen Satz, in dem es um die Hauptstadt Deutschlands geht, meist ähnlich, wie „... *Berlin, als Hauptstadt der Bundesrepublik ...*“. Diese Recherchemethode ist also auch als durchaus erfolgreich zu betrachten, man muß nur bei der *Frage*-Stellung darauf achten, daß möglichst wenig Stopfwords und möglichst viele relevante Stich- und Schlagworte in der Frage enthalten sind.

2.2 Profi-Suche

Advanced Search bei AltaVista bringt einige erhebliche Veränderungen gegenüber der einfachen Suche. In diesem Modus wird die volle Breite der bool'schen Operatoren, NEAR und auch Klammerung verfügbar. Des weiteren gibt es auch ein einstellbares Ranking und die Möglichkeit die Treffer zeitlich zu begrenzen. *Sehr wichtig*: Es findet kein automatisches Ranking statt. Der Benutzer muß die Kriterien in dem dafür vorgesehen Feld angeben.

2.3 Firmengeheimnisse der Suchmaschinen

Die hier beschriebenen ‚*Verhaltensweisen*‘ der Suchmaschinen entstammen der Hilfestellung und Beobachtungen mittels Proberecherchen. Es ist aber manchmal zu beobachten, daß die Suchmaschine nicht präzise wie erwartet reagieren, sondern nur fast (siehe auch Groß- und Kleinschreibung und Umlaute/Accents). Leider ist nicht genau zu erfahren, welche Methoden, welche Algorithmen den Suchmaschinen zugrunde liegen, weil diese Informationen als Firmengeheimnisse betrachtet werden.

3 Beurteilungskriterien

Zuerst sollten einige Kriterien aufgestellt werden, die die objektive Beurteilung der einzelnen Suchdienste erlauben. Unerfahrene Benutzer werden wegen mangelnder Routine und fehlender Retrievalkenntnisse ihre Suchanfragen einfach stellen. Für sie wird es wichtig sein, daß sich die Möglichkeiten zur Einschränkung der Treffermenge einfach, gut erklärt, gleich neben der Suche finden. Durch unpräzise Suchanfragen werden sie wohl wesentlich mehr Treffer als Profis bekommen. Deswegen ist es sehr wichtig, daß die Treffer in der Anzeigeliiste nach Relevanz möglichst gut ‚*gerankt*‘ werden. Man kann also sagen, daß für einfache Sucher eine kurze,

deutliche und gut erreichbare Hilfestellung, gutes ‚*Ranking*‘ und ausreichende Größe des Bestandes die wichtigsten Kriterien sein werden.

Profisucher hingegen geben sich mit Ergebnissen, die mehrere tausend Dokumente umfassen, nicht zufrieden. Sie werden ihre Suche weiter präzisieren wollen. Jetzt kommt es darauf an, daß eine Suchmaschine viele Einstellmöglichkeiten also ein gutes Retrieval bietet.

Daher ergeben sich folgende Kriterien für die Beurteilung:

1. *Art der Erschließung der Dokumente*: Werden die Dokumente vom Suchroboter Volltext erschlossen, oder nur einige wichtige Felder? Werden Dokumente per Registrierung dem Bestand hinzugefügt, muß der Benutzer die Eingliederung in die Systematikgruppen selbst vornehmen?
2. *Einfache und Profisuche*: Was bietet die Profisuche gegenüber der einfachen?
3. *Größe des Katalogs*: Wieviele Dokumente befinden sich in der Datenbank?
4. *Bool'sche Logik*: Welche bool'schen Operatoren können benutzt werden? Ist eine individuelle Klammerung möglich?
5. *Trunkierung*: Sind die Möglichkeiten für Links- oder gar für Innentrunkierung geschaffen? Oder erfolgt die Trunkierung automatisch.
6. *Besteht die Möglichkeit eine Phrasensuche durchzuführen?*
7. *Die Groß/Kleinschreibung*: Findet eine Unterscheidung statt?
8. *Feldsuche*: Kann man nach Inhalten von Feldern, die elektronischen Bibliothekskatalogen ähnlich sind, suchen (Titel, URL, Domain, usw.)?
9. *Kontextsuche*: Suche nach aufeinanderfolgenden Schlagworten?
10. *Suche nach verschiedenen Medientypen*: Bilder, Sound, Videos, VRLM
11. *Ranking*: kann man Einstellungen treffen, die die Reihenfolge der angezeigten Treffer beeinflusst? Oder findet ein automatisches Ranking statt?
12. *Hilfestellung*: Gibt es eine ausreichende Hilfestellung, die knapp, aber ausführlich genug ist? Wo ist diese angebracht, erscheinen Tips neben den Suchfeldern?
13. *Die Benutzerfreundlichkeit*: Es ist schwierig zu definieren, was die Benutzerfreundlichkeit einer Suchmaschine ausmacht. Es sind allerdings einige Tendenzen zu nennen, die dazu beisteuern: Intuitive Benutzung (wenig EDV-Kenntnisse erforderlich, die Benutzung erfolgt nach ‚*Gefühl*‘) Maus-Bedienung oder Text-Retrieval, eine Bildschirmästhetik (die selbstverständlich nur subjektiv beurteilt werden kann).

Diese Erwartungen wurden von den verschiedenen Suchmaschinen verschieden, auf unterschiedliche Arten und unterschiedlich gut oder schlecht erfüllt.

4 Einzelne Suchmaschinen

4.1 AltaVista¹

AltaVista ist eine der größten Volltextdatenbanken mit Schlagwortsuche. Ihr Bestand liegt momentan (Anfang

1 <http://www.altavista.com>

Nov. 1998) nach eigenen Angaben des Betreibers bei ca. 140 Mill. Web-Seiten und 16 Tausend Newsgroups. Sie wickelt pro Tag ca. 32 Millionen Suchanfragen ab.

Da die Funktionsweise von AltaVista bereits dargestellt wurde, gebe ich hier nur noch eine kurze Zusammenfassung über die Potentiale dieser Suchmaschine:

Bei ‚*simple search*‘ stehen den Benutzern einfache Schlagwortsuche, Phrasensuche, die Benutzung der +/– Zeichen, Rechts- und Innentrunkierung (nach dem 3. Zeichen), sowie Feldsuche zur Verfügung. In diesem Modus erfolgt auch ein automatisches Ranking nach den eingegebenen Suchthermen.

Im Modus ‚*advanced search*‘ werden die Suchmöglichkeiten um die bool'schen Operatoren und den NEAR-Operator, sowie individuelle Klammerung erweitert. In diesem Modus findet auch kein automatisches Ranking mehr statt; der Benutzer ist gezwungen, selbst die Kriterien des Rankings zu bestimmen.

Zur Nachbearbeitung der Recherchen steht das ‚*refine*‘ zur Verfügung. Hier werden die Begriffe aufgelistet, die in den Trefferdokumenten besonders häufig anzutreffen sind. Hier ist auch die grafische Darstellung zu beachten, die Zusammenhänge zwischen diesen häufigen Begriffen gut verdeutlicht.

AltaVista bietet auch einen sehr einfach gegliederten systematischen Katalog und auch andere Dienste, u.a. Suche nach Newsgroups-Beiträgen, Fotos, Leuten, Firmen, Karten sowie Benutzerprofile und ein automatisches Übersetzungsservice für einige westeuropäische Sprachen.

Beurteilung: Die Hilfestellung ist kurz und verständlich. Das Gesamtkonzept von AltaVista stützt sich auf Textretrieval. Dies ist zwar heutzutage nicht mehr ganz in, aber es ist immer noch sehr verständlich und sorgt für eine hohe Präzision.

Diese Suchmaschine überzeugt mit einem großen Bestand, sehr guten Recherchemöglichkeiten und schnellen Antwortzeiten. Mit *refine* bietet AltaVista die Möglichkeit große Treffermengen bequem unter Kontrolle zu kriegen, und schafft auch den Ausbruch aus der Textretrieval-Umgebung.

AltaVista bietet unter den besprochenen Suchmaschinen die besten Möglichkeiten, eine Suchfrage präzise zu formulieren, und diese später auf das relevanteste an Treffern zu reduzieren. Aus diesem Grund ist sie bis heute für mich erste Wahl.

4.2 Fireball

Fireball hatte im April nach eigenen Angabe rund 6,5 Millionen deutsche WWW-Dokumente im Katalog. Fireball ist zwar eine Schlagwort-Suchmaschine, setzt aber auch auf systematische Verzeichnisse und die grafische Darstellung.

Sie bietet auch eine ganze Menge von systematisch aufbereiteten Katalogen an, die alle mit der Hauptsystematik auf der Homepage vertreten sind. Hierunter sind *Trends*, *Guides*, ‚*Themen täglich frisch*‘ und *Surftips* zu finden. Es ist hier über *Paperball* auch möglich in den aktuellen Nachrichten der Zeitungen zu recherchieren. Paperball ist über eine kurze Systematik aus 6 Gruppen erschlossen, und bietet auch die Möglichkeit eine Schlagwortrecherche durchzuführen, bei der nach Zeitung und Rubrik eingeschränkt werden kann.

Da Fireball mit AltaVista kooperiert sind die meisten Suchfunktionen identisch. Die *Detailsuche* von Fireball ist ein wenig anders als ‚*Advanced Search*‘ bei AltaVista. Fireball setzt hier mehr auf eine grafische Darstellung, aber die bekannten AltaVista Optionen sind schnell wiederzufinden.

Beurteilung: Fireball ist zwar eine schlanke Version von AltaVista, bietet aber durch nützliche zusätzliche Dienste einen sehr guten und verständlichen Einstieg in die deutschsprachige Recherche.

4.3 DINO², Lotse³ und Branchenbuch⁴

DINO erschlägt den Benutzer gleich auf der Homepage mit einer Fülle von Diensten aber mit kaum einer Suchmöglichkeit. Angefangen vom Webkatalog mit 25 Hauptsachgruppen über aktuelle Meldungen, regionale Suchmöglichkeiten bis hin zu den Suchmaschinen. Der Bestand wird per Registrierung durch Web-Autoren erweitert.

Katalog: Der Webkatalog hat 6400 Themenseiten (Sachgruppen) 25 Oberkategorien (Hauptsachgruppen) untergeordnet. Die Katalogseite ist ewig lang und nur wenig übersichtlich, denn unter den Hauptsachgruppen sind auch einige (zum Glück nicht alle) der 6400 Sachgruppen angezeigt. Die Sachgruppen sind auch alphabetisch erschlossen.

Innerhalb der Hierarchie besteht dann die Möglichkeit eine einfache Suche im gesamten Katalog oder in der aktuellen Hauptsachgruppe durchführen zu lassen. Dies vereinfacht die Navigation zwischen den zahlreichen Sachgruppen entscheidend.

Die DINO-Suchmaschinen: Hier findet man drei Möglichkeiten der Schlagwortsuche. Bei DINO kann man in einem redaktionell (!) bearbeiteten Katalog mit ca. 170 Tausend Webseiten suchen. Dieser Suchdienst kann einfache Schlagwortsuche, Rechtstrunkierung, die Groß- und Kleinschreibung wird beachtet, und auch die bool'schen Operatoren (AND, OR und NOT) sind erlaubt. Worte, die mit einem einfachen Leerzeichen verbunden sind, werden, wie mit AND-verknüpfte behandelt. Es wird *keine Volltextsuche* durchgeführt! Die Trefferliste zeigt nicht nur die einzelnen Trefferdokumente an, sondern auch die Sachgruppen im systematischen Katalog.

Diese Suchmöglichkeit ist im Vergleich zu Suchmaschinen sehr dürftig. Es gibt keine Phrasensuche, keine +/– Benutzung, kein einstellbares Ranking, keine Kontextsuche, keine Möglichkeit zur Nachbearbeitung. Auch die Hilfestellung ist sehr kurz gefaßt.

Lotse Suche: Die Defizite der DINO-Suche versucht dieser Suchdienst über die Kooperation mit der deutschsprachigen Volltextsuchmaschine *Lotse* auszugleichen. Bei DINO ist der ‚*richtige*‘ Schlagwortsuchdienst Lotse und bei Lotse ist der Katalog DINO.

Lotse bietet die volle Breite der bool'schen Operatoren samt NEAR und individueller Klammerung der Begriffe. Zu diesem Suchdienst gibt es auch eine ausreichende Hilfestellung, die sogar über das Ranking berichtet.

Bei Lotse fehlt insbesondere die Phrasensuche, und einige Benutzer könnten auch die +/– Benutzung vermisse-

2 <http://www.dino-online.de/>

3 <http://www.lotse.de/>

4 <http://www.branchenbuch.com/>

sen. Es werden aber auch interessante weitere Suchdienste angeboten: *Film.de* *Donnerwetter.de* *Urlaub.de* *Finanzen.de* *Auto.de* *Stellenmagazin.de*. Leider konnte ich nirgends einen Verweis auf die Größe der Datenbank oder auf die Benutzungsdaten finden.

Das Branchenbuch: Dieser Suchdienst entstand in Kooperation mit dem ‚Großen Deutschen Branchenbuch im Internet‘. Es ermöglicht die Recherche in mehr als drei Millionen deutschen Firmenadressen und gliedert sich in drei Suchmodi: In der einfachen Suche kann der Benutzer eine Suche nach Stichworten oder Branchen starten. Bei der Detailsuche kann man eine Art Feldsuche durchführen (Name, Straße, PLZ, Ort, usw.). In der Themensuche ist ein systematischer Zugang realisiert, der ähnlich DINO's Katalogseite auch die Untergruppen – allerdings hier in Dropdown-Menüs versteckt – enthält. Neben der sehr ansprechenden grafischen Aufbereitung wäre auch anzumerken, daß neben den Suchfeldern die Tips für die Benutzung fehlen.

Beurteilung: die Kooperation mit anderen Firmen tut DINO besonders gut, und läßt die vielen Schwächen des Katalogs und der Suchmaschine fast vergessen, aber es ist überall zu sehen, daß die besseren technischen Lösungen und schönere grafische Darstellungen erst in der Zusammenarbeit mit Lotse und dem Branchenbuch entstanden. Alles berücksichtigt kann man sagen, daß der DINO-Zusammenschluß auf dem Wege ist eine vernünftiger, umfassender Suchdienst zu werden, heute aber noch mit den Defiziten des Suchretrievals und einiger Konzeptlosigkeit zu kämpfen hat.

4.4 HotBot⁵

HotBot ist einer der größten Volltext-Schlagwortkataloge im Internet überhaupt. Sie ist eine Besonderheit unter den besprochenen Suchdiensten. Sie arbeitet mit einer Oberfläche, die sehr Maus- und grafisch orientiert ist. Mit Dropdown-Menüs und *Checkboxes* werden schnell einfache Schlagwortsuche, Phrasensuche, Feldsuche, eine Art bool'sche Suche, Mediensuche, zeitliche und sprachliche Begrenzung der Treffermenge, sowie verschiedene Formate für die Trefferdarstellung realisiert. Die Suchmaske und die Suchoptionen von HotBot sind so aufgebaut, daß die Übergänge zwischen einfacher Suche und professioneller Suche fließend sind.

Schauen wir uns die verschiedensten Einstellmöglichkeiten an: In der ersten Eingabezeile kann man eine Suchfrage aus mehreren Suchbegriffen formulieren. Die Beziehung dieser Begriffe zueinander wird durch das Dropdown-Menü darunter bei den ‚*Search Options*‘ definiert. ‚*all the words*‘ ist eine AND-Verknüpfung, ‚*any of the words*‘ eine OR-Verknüpfung, ‚*exact phrase*‘ die Phrasensuche, ‚*page title*‘, ‚*the person*‘, ‚*link to this URL*‘ sind Suchfelder, und schließlich ‚*Boolean phrase*‘ ein bool'scher Ausdruck, den der Benutzer frei formulieren kann. Gleich darunter kann man eine zeitliche und eine sprachliche Begrenzung der Treffermenge vornehmen. Unter all diesen Optionen gelangt der Benutzer zu der ‚*SuperSearch*‘, weiter darunter sind die verschiedenen Dienste (*Usenet*, *Yellow Pages*, e-Mail-Adressensuche, usw.) zu erreichen. Auf der rechten Seite ist ein einfacher systematischer Katalog zu sehen, der eigentlich nur zu populären Themen einen Einstieg bietet.

Bei der *SuperSearch* präsentiert sich HotBot den Benutzern von ihrer besten Seite, eine ganze Menge von gut

formulierten, verständlichen Suchoptionen und Feldern, die mit dem grafischen + Zeichen um weitere ergänzt werden können. Auch die bool'schen Operatoren sind in ihrer ausformulierten Form anzutreffen, ‚*must contain*‘, ‚*should contain*‘ und ‚*must not contain*‘ sprechen für sich. Die Optionsfelder sind sehr gut beschriftet und daneben erklärt.

Umlaute und Klein- und Großschreibung werden natürlich beachtet. Trunkierung, Klammerung und Kontextsuche sind allerdings bei dieser Suchmaschine nicht möglich, aber dieses Manko machen all die anderen Vorteile wieder wett.

Beurteilung: Das Glanzstück von HotBot ist die grafische Schnittstelle zwischen dem Textretrieval und dem Benutzer. Sie ist eine sehr gut durchdachte Suchmaschine, die neben intuitiver Benutzung auch die Optionen zu professionellen Textretrieval bietet. Die Hilfestellung ist kurz, ausführlich und verständlich genug.

HotBot ist für Anfänger als auch für Profis sehr empfehlenswert.

4.5 Lycos.de⁶

Nach eigenen Angaben hat Lycos (international) in ihrer Datenbank mehr als 100 Millionen URLs, sie zählt also zu den großen. Lycos ist aber wahrscheinlich keine Volltext-Datenbank. Allem Anscheine nach werden bei Lycos die Formalen Daten und die ersten 20 Zeilen des Dokumentes erfaßt. Sie meldet sich mit einer gut gefüllten Oberfläche. Auf der rechten Seite sind die Lycos-Dienste zu finden, in der Mitte sind die verschiedenen Suchmöglichkeiten auf Reitern angezeigt (einfache und Profisuche, Suche nach Nachrichten, e-Mail-Adressen, Sounds und Bildern, Firmen und in der Top 1000 Datenbank). Darunter wird aber auch ein systematischer Einstieg in die Datenbank der vorrecherchierten Themenbereiche.

Der Benutzer trägt die Suchbegriffe in die Eingabezeile ein, dann wird ein Katalog gewählt (*gesamtes Internet*, *deutschsprachiger Raum*, *Bücher-Katalog*, *Sounds*, *Bilder* und das *Wörterbuch*) darunter kann man einstellen welche Sprache die Trefferdokumente haben sollen, und zuletzt neben den beiden Optionen kann man die Verbindung der eingetragenen Suchbegriffe bestimmen: *Alle Wörter* bedeutet, daß die Begriffe mit AND, *Eins der Wörter*, daß diese mit OR verknüpft werden, und *Genauer Text* steht hier für die Phrasensuche. Bei dieser einfachen Suche können auch die bekannten +/- Zeichen benutzt werden.

Die Profi-Suche bietet eine drei geteilte Suchmaske. Im oberen Teil kann man in einer Eingabezeile die Suchfrage formulieren. Es sind die bekannten bool'schen Operatoren erlaubt, Klammersetzung, und eine ganze Menge (!) an Kontext-Operatoren, die dadurch eine sehr feine Einstellung der Suche erlauben, daß die Entfernung und die Reihenfolge der Kontextsuchbegriffe in den Trefferdokumenten eingestellt wird. Es können die Verknüpfungsart, der Katalog und die Sprache eingestellt werden. Der Benutzer kann auch angeben in welchem Bereich der Dokumente die Suche durchgeführt werden soll (*im ganzen Dokument*, *im Titel*, *im URL*) und auch das Domain der Treffer kann angegeben wer-

5 <http://www.hotbot.com/index.html>

6 <http://www.lycos.de/>

den. Darunter befindet sich der Bereich, in dem die Optionen für die Sortierung (Ranking) untergebracht wurden. Eine Tabelle erleichtert die Auswahl. Ganz unten kann man die Anzeige der Treffer einstellen, die Länge der Trefferlisten bzw. nach welchem Kriterium sie sortiert werden: nach *Domäne* (Alle Treffer, die sich auf einem Server befinden werden untereinander gelistet) oder nach *Relevanz* (die Treffer zuerst angezeigt, die auch am wahrscheinlichsten relevant sind).

In der Trefferliste findet sich neben jedem Treffer ein eigener Verweis auf *Ähnliche Ergebnisse*, dies bietet eine Möglichkeit zur Nachbearbeitung einer bereits durchgeführten Recherche. Bei dieser Nachbearbeitung bleiben dem Benutzer leider keine Einstellmöglichkeiten. Die Nachrecherche ist davon abhängig, zu welchem Treffer ähnliche Dokumente gefunden werden sollten, alles weiter übernimmt Lycos.

Beurteilung: Lycos präsentiert für Anfänger eine mittelmäßig verständliche Suchoberfläche, und für Profis ein Suchretrieval, das sehr ausgefeilte Mittel enthält. Es gibt zu allen Bereichen verständliche kurze Erklärungen und auch Hilfestellungen, die aber nicht alle Aspekte ansprechen. Die gesamte Suchmaschine ist von einem gewissen Chaos durchwoben. Es gibt immer wieder Bereiche, in denen der Benutzer Optionen einstellen kann, die durch Optionen anderer Bereich zum größten Teil überlappt werden. Die Suche *im deutschsprachigen Bereich* und eine Trefferliste mit nur *deutschsprachigen* Dokumenten bedeuten in den meisten Fällen dasselbe. Es finden sich aber auch noch mehr solche Unstimmigkeiten, bis hin zu toten Verweisen der Suchoberfläche und zum Teil englischsprachiger Hilfestellung in der ansonsten vollkommen deutschen Suchmaske. Die Suchmöglichkeiten sind bei dieser Suchmaschine zahlreich, doch chaotisch organisiert, deswegen, bleibt die Benutzung dieser Suchmaschine gewöhnungsbedürftig.

4.6 Yahoo!.de⁷

Yahoo! bietet einen systematischen Katalog mit der Möglichkeit einer Schlagwortrecherche im Bestand und unter den Sachgruppen. Bestandserweiterung erfolgt über Registrierung der neuen Web-Seiten durch die Autoren.

Die Suchmaschine meldet sich mit einer sehr einfachen Oberfläche. Eine Eingabezeile für die Formulierung der Suchfrage ist vorhanden, darunter wird aber der wahre systematische Katalog von Yahoo! sichtbar. Eine einfache Systematik, die 14 Gruppen umfaßt. Neben der Systematik sind auch noch aktuelle Nachrichten, Meldungen und Dienste von Yahoo! aufgelistet.

In der Eingabezeile können +/- Zeichen, Trunkierung, Phrasensuche und eine begrenzte Feldsuche (*t...* für Begriffe im Titel, und *u:...* für Begriffe im URL) durchgeführt werden. Die Groß- und Kleinschreibung wird dabei nicht beachtet.

Die Anzeige der Treffer besteht aus zwei wesentlichen Teilen: zuerst werden die gefundenen Kategorien aufgelistet, dann auch die einzelnen Web-Seiten, denen aber immer die Sachgruppe vorangeht, in der diese zu finden waren. Geht der Benutzer zu einer der Sachgruppen, so erhält er auch die Möglichkeit einer Nachrecherche in der Sachgruppe, oder einer neuen Recherche in der gesamten Datenbank von Yahoo!.de oder dem internationalen Yahoo!-Katalog durchzuführen.

Bei der Navigation zwischen den Sachgruppen fällt auf, daß nach der obersten Ebene die Anzahl der Untergruppen und Einträge neben den Sachgruppen aufgeführt wird, und auch ob neue Einträge in der Gruppe enthalten sind.

Beurteilung: Das Retrieval von Yahoo!'s Suchmaschine ist zu einfach realisiert. Es fehlen wichtige Optionen, wie bool'sche Operatoren, Klammerung, Unterscheidung der Groß- und Kleinschreibung. Die Hilfestellung ist ausreichend. Der Katalog selbst ist gut geordnet übersichtlich. Es sind Yahoo!'s zahlreiche Dienste lobend zu erwähnen, unter denen auch andere deutsche Suchdienste anzutreffen sind.

4.7 Web.de⁸

Web.de behauptet gleich auf der Homepage von sich, Deutschlands umfangreichster Web-Katalog zu sein und auf den Folgeseiten die redaktionell bearbeiteten Einträge gut strukturiert (in 16 Rubriken) und aktuell anzubieten. Auch bei diesem systematischen Katalog werden die Benutzer aufgefordert den Katalog durch die Registrierung neuer Angebote zu erweitern.

In der einfachen Suche findet man nur wenige Möglichkeiten die Suchfrage präzise zu formulieren. Es ist möglich nach Phrasen zu suchen und mit einem Pluszeichen kann man Suchbegriffe verbinden, wobei anscheinend eine AND-Verknüpfung realisiert wird. Bei der Suche findet eine automatische Rechts- und Linkstrunkierung statt. In der Trefferliste erscheint eine Möglichkeit der Nachrecherche in den anderen Katalogen von Web.de, danach die gefundenen Kategorien zur Suchanfrage, und dann die einzelnen gefundenen Treffer.

Bei den *Optionen* findet man eine Art Profi-Suche. Hier hat der Benutzer die Möglichkeit die Verknüpfungsart der Suchbegriffe auszuwählen, wobei nur AND und OR zur Verfügung stehen, sowie die Beachtung der Groß- und Kleinschreibung zu erzwingen und die automatische Trunkierung auf eine automatische Rechtstrunkierung zu beschränken.

Beurteilung: Anfang November 1998 hatte Web.de nach eigenen Angaben ,161 782 deutsche Homepages', und soll der umfangreichste deutsche Web-Katalog sein, wobei DINO von ca. 170 000 Einträgen spricht. Der Web-Katalog von Web.de ist gut strukturiert und ansprechend auf einen Blick übersichtlich präsentiert. Das Suchretrieval dagegen läßt sehr viel zu wünschen übrig. Nicht nur, das viele Suchmöglichkeiten fehlen, sondern auch bei den Optionen finden sich keine nennenswerten Neuerungen gegenüber der einfachen Suche. Bei diesem Suchretrieval kann allenfalls von einem Spielzeug die Rede sein.

4.8 Eule.de⁹

Die Homepage von Eule ist Web.de sehr ähnlich doch etwas einfacher von der grafischen Aufbereitung her. Auf der linken Seite befinden sich die Dienste von Eule und die Informationen über sie. Auf der anderen Seite sehen die Benutzer die Systematik von Yahoo!.de und auch jeweils eine kurze Auslese daraus. Über der Sy-

7 <http://www.yahoo.de/>

8 <http://web.de/>

9 <http://www.eule.de>

stematik befinden sich die Suchtips und die Spezial-suchdienste vor allem das Suchfeld mit den Optionen zur Anzeige der Trefferliste.

In der einfachen Suche stehen dem Benutzer einfache Schlagwortsuche und die Benutzung der +/- Zeichen zur Verfügung. Die Hilfestellung gibt einen kurzen Hinweis, daß die deutschen Umlaute sowohl als *ä* als auch als *ae* richtig recherchiert werden. In der Profi-Suche kommt noch die Möglichkeit zur Einstellung der Sortierkriterien hinzu.

Beurteilung: Die Suchmaschine selbst hat so wenige Einstellmöglichkeiten, daß sie damit die schlechteste unter den besprochenen ist. Auch die Profi-Suche bringt keine nennenswerten Neuerungen in das Suchretrieval. Bei einer Stichprobe ergab sich, daß AltaVista unter den deutschsprachigen Dokumenten ca. 9 mal so viel und Fibreball ca. 10 mal so viele Treffer erzielt. Da der systematisch Katalog ja auch von Yahoo! ist, kann man sagen, daß diese Suchmaschine auch nur als Spielzeug taugt.

4.9 Eine Suchmaschine fürs Leben

Es gibt wohl nur wenige Internetbenutzer die bei ihren Recherchen mehrere Suchdienste konsultieren. Anscheinend ist es eine für eine längere Zeit (fast fürs Leben), wenn man sich aus dem großen Angebot einen Suchdienst auswählt. Steht man (ob Fachmann oder als Privatperson) vor dieser Wahl, würde ich empfehlen, daß man zuerst mehrere Suchdienste anschaut, auch deren Hilfestellungen und sich dann für das sympathischste entscheidet. Diese wird nämlich aller größter Wahrscheinlichkeit der Suchdienst sein, mit dem der Benutzer am besten umgehen kann. Hier noch eine Liste von größeren, interessanteren Suchdiensten, die hier nicht behandelt wurden. International: Excite¹⁰, Infoseek¹¹, Magellan¹², OpenText¹³; deutsche: Excite.de¹⁴, Kolibri¹⁵.

5 Suche nach anderen Informationen

Außer der Suche nach HTML-Dokumenten kann man nach vielen anderen Arten von Informationen recherchieren. Diese Recherchemöglichkeiten werden in der Regel von den Suchmaschinen selbst angeboten. Es ist zu empfehlen, bei den größeren Suchmaschinen nachzuschauen, welche zusätzliche Dienste angeboten werden. Hier möchte ich nur einige erwähnen:

Suche nach Newsgroups-Beiträgen: Diese Funktion wird von den meisten besprochenen Suchdiensten unterstützt. AltaVista: *Usenet*, HotBot: *Usenet*,

Suche nach e-Mail-Adressen: Die Suche nach E-Mail-Adressen gestaltet sich ziemlich schwierig dadurch, daß es im Internet keine zentrale Registrierung gibt. Solche Daten werden von Organisationen gesammelt und als eine Art öffentliches Telefonbuch im Internet zugänglich gemacht. Auch Suchmaschinen betreiben die Sammlung von E-Mail-Adressen anhand von Newsgroupsbeiträgen. Automatisch erfaßt werden also nur die Adressen von Leuten, die in Newsgroups aktiv mitmachen. AltaVista: *People Finder*, HotBot: *Email Addresses*, Yahoo.de: *E-Mail-Adressen*, Lycos.de: *eMail*;

Suche nach Firmeninformationen: Im Regelfall sind es separate Datenbanken oder gar andere Anbieter per Kooperation: HotBot: *Yellow Pages*, Yahoo.de: *Branchenbuch*, Lycos.de: *Firmen*, DINO: *Branchenbuch*, Al-

taVista: *Yellow Pages*, und auch am Anfang einer jeder Trefferliste findet sich ein kleiner Verweis: »Suchbegriff“ *Official company or product home page by RealName*.


Quellennachweis

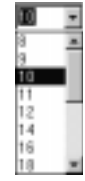
- Die kleine Suchfibel Intro ... <http://www.suchfibel.de/>
- Diese Zusammenstellung ist besonders empfehlenswert, weil sie sehr gut geordnet ist, und weil dort auch Vergleiche bezüglich der Features und des Bediensyntaxes zwischen den Suchmaschinen zu finden sind.
- Goldenen Regeln für schlechtes HTML: <http://www.karzauninkat.com/Goldhtml/>
- Suchmöglichkeiten im Internet http://www.ib.hu-berlin.de/inf/i_suche.htm
- Des weiteren hatte ich die Suchhilfen, Beispiele, FAQs und sonstigen Angaben der einzelnen Suchmaschinen als Quelle. Vieles davon, was ich hier beschrieben hatte stützt sich auf Beobachtungen und Proberecherchen

Fachglossar

Geschlecht von Suchmaschinen: Da ich nicht entscheiden konnte, ob Yahoo! (und auch andere Suchmaschinen) männlich, weiblich oder neutral sind, habe ich mich entschieden ähnlich den Seefahrern ‚meine‘ Gefährten als weibliche Wesen zu betrachten.

Boolsche Operatoren: Aus der Mathematik (Mengenlehre/Logik) bekannten Operatoren zum Verknüpfen von Mengen: *and*: Schnittmengenbildung, *or*: Vereinigungsmengenbildung, *not*: Ausschließen.

checkbox: Ein kleines Kästchen zum An- und Auskreuzen: 

Dropdown-Menü: Ein Menü zum ‚Runterklappen‘, z.B.: das bekannte kleine Menü zum Einstellen der Schriftgröße: 

Links: (auch Hyperlinks genannt) sind aktive Verknüpfungen, das heißt aktive Verweise auf andere Quellen. Beim Anklicken werden die Bezüge automatisch hergestellt. Diese Art der Verweise wird auch interaktiv genannt. Solche Verweise findet man bei interaktiven CD-ROMs und vermehrt im Internet, weil diese Medien verhältnismäßig schnell bei der Ausführung dieser interaktiven Verweise sind.

Schlag- oder Stichwortkatalog: Da Schlagwortkataloge Dokumente oft in Volltext aufnehmen, kann man nicht mehr über Schlag- oder Stichworte reden. Diese Unterteilung spielt in dem Zusammenhang eine sehr untergeordnete Rolle. Schlagworte können ja auch gleichzeitig Stichworte sein, aus diesem Grund werde ich weiterhin nur über Schlagworte sprechen.

Tag: Formatierungselemente in HTML-Dokumenten, die sich durch spitzen Klammern abheben und gut zu indizieren sind. Bsp.: `` ein Bild wird eingefügt

„tote Links“: unbrauchbare Hyperverweise. Meistens erfolgt eine Fehlermeldung, daß entweder der entsprechende Server oder die entsprechende Seite nicht gefunden wird.

URL: [Universal Resource Locator] Adresse eines Internet-Dokuments.

Anschrift des Autors:

Ferenc Becker
Schottenstr. 73
D-78462 Konstanz

10 <http://www.excite.com/>

11 <http://www2.infoseek.com/>

12 <http://www.mckinley.com/>

13 <http://pinstripe.opentext.com/>

14 <http://www.excite.de/>

15 <http://www.kolibri.de/>