Abfragesprachen und erweiterte Funktionen von WWW-Suchmaschinen

Dirk Lewandowski, Düsseldorf

Der Artikel stellt die erweiterten Suchmöglichkeiten in den wichtigsten Suchmaschinen vor. Dabei handelt es sich um Google, Alta Vista, Hotbot (Inktomi), Teoma und Fireball. Die Qualität von Suchmaschinen wird in der Regel durch Retrievaltests verglichen. Vor dem Hintergrund professioneller Ansprüche an die Recherchefähigkeiten der Suchmaschinen müssen aber zusätzlich auch deren Abfragemöglichkeiten betrachtet werden. Nur ausgereifte Abfragesprachen erlauben komplexe Suchanfragen, wie sie im professionellen Umfeld gestellt werden.

Query languages and extended functions of www search engines

This article presents the advanced search possibilities in the most important search engines: Google, Alta Vista, Hotbot (Inktomi), Teoma and Fireball. The quality of search engines is usually compared through retrieval tests. Against the background of professional demands on the abilities of search engines, however, their query languages must be regarded additionally. Only perfected query languages permit complex retrieval as it is placed in a professional sphere.

Einleitung

Als beliebte und weit verbreitete Angebote werden Suchmaschinen vor allem von den „gewöhnlichen Nutzern" des Internet benutzt. Bisher vorliegende Untersuchungen haben gezeigt, dass diese nicht dazu bereit sind, sich mit Operatoren und erweiterten Suchformularen auseinander zu setzen (ausführlich in Silverstein/Henzinger/Marais/Moricz 1998). Die Suchmaschinen haben darauf reagiert, indem erweiterte Suchformulare nicht mehr oder nur noch eingeschränkt weiterentwickelt werden. Während die „klassische" Suchmaschine *Alta Vista* sich bei ihrem Start im Jahr 1995 noch an den Möglichkeiten professioneller Retrievalsysteme aus dem Bereich der Hosts orientierte, stellen neuere Suchmaschinen oft nur wenige Kommandos zur Verfügung.

Dies lenkt auf die Frage hin, in wie weit heutige Suchmaschinen den Anforderungen der Information Professionals gerecht werden, die letztlich ja auch in Konkurrenz zu den Alltagsnutzern stehen, die von sich behaupten, „in Google alles finden zu können."

In diesem Aufsatz werden die Kommandos und Einschränkungsmöglichkeiten der wichtigsten Suchmaschinen erläutert und es wird auf ihre Besonderheiten hingewiesen. Dabei werden nur die Möglichkeiten der Recherche nach Texten behandelt; auf die Suche nach Multimedia-Inhalten wird nicht eingegangen.

Retrievaltests

Üblicherweise werden Suchmaschinen durch Retrievaltests miteinander verglichen (vgl. u.a. Griesbaum/Rittberger/Bekavac 2002, Veritest 2003). Dabei werden die gleichen Suchanfragen an unterschiedliche Suchmaschinen gestellt und die zurückgegebenen Ergebnisse verglichen. In der Regel werden dazu die jeweiligen Top-Platzierungen der einzelnen Suchmaschinen ausgewertet (in der Regel die ersten zehn oder 20 Treffer).

Bei den Retrievaltests können allerdings nur relativ unpräzise Suchanfragen gestellt werden, weil sich die Suchsprachen der einzelnen Suchmaschinen zu weit voneinander unterscheiden, um die Ergebnisse präziser Anfragen vergleichbar zu machen. Auch für den Nutzer ausgesprochen hilfreiche Methoden, um die Suchanfrage nach dem ersten Ansehen der Trefferliste weiter einzuschränken, können bei Retrievaltests nicht berücksichtigt werden. Es ist also festzustellen, dass Retrievaltests den Möglichkeiten einzelner Suchmaschinen nicht gerecht werden können. Vielmehr müssen die Tests sich mit relativ einfachen Suchanfragen begnügen. Ihre Ergebnisse sind daher eher für die Anfragen der gewöhnlichen Nutzer aussagekräftig; dass eine Suchmaschine einfache Anfragen in hoher Qualität beantworten kann, macht sie jedoch noch nicht zur ersten Wahl des Information Professionals.

Insbesondere die Suchmaschine *Google* erfreut sich größter Beliebtheit und belegt in (vor allem populärwissenschaftlichen) Retrievaltests in der Regel den ersten Platz. Offensichtlich gelingt es dieser Suchmaschine, einfache Suchanfragen mit hoher Präzision zu beantworten. Oder – um der Anlage der Retrievaltest gerecht zu werden – es gelingt ihr, irrelevante Dokumente aus der Top-10 bzw. Top-20 herauszuhalten. In den vorliegenden Retrievaltests sind allerdings die aktuellen Entwicklungen hin zu einem massiven Spamming des *Google*-Index noch nicht berücksichtigt. (Karzaunikat 2003)

Interessanterweise ist es allerdings gerade die Suchmaschine *Google*, die wenig Wert auf präzise Abfragemöglichkeiten durch die Suchsprache oder erweiterte Suchformulare legt. Auch bei anderen Suchmaschinen lässt sich feststellen, dass die Abfragesprachen oft zu wünschen übrig lassen. Als Argument wird seitens der Suchmaschinen-Betreiber angeführt, dass diese erweiterten Funktionen nur selten genutzt würden. Allerdings gibt es keine Erhebungen zu der Frage, ob diese nicht etwa von einer kleinen Gruppe von Nutzern intensiv genutzt werden.

Untersuchte Suchmaschinen

Für diese Untersuchung wurden die Suchmaschinen *Google, All The Web, Alta Vista, Hotbot* (Inktomi), *Teoma* und *Fireball* ausgewählt. Hierbei handelt es sich um die Suchmaschinen mit den weltweit größten Indexen (vgl. Sullivan 2003) bzw. im Fall von Fireball um die führende deutsche Suchmaschine mit eigenem Index. Portalangebote wie *Yahoo!* oder *T-Online* wurden nicht berücksichtigt, weil diese ihre Suchergebnisse in nahezu allen Fällen von einer der genannten Suchmaschinen beziehen. (vgl. Clay 2002; für den deutschen Markt vgl. Suchfibel 2003)

Hotbot wurde ausgewählt, weil diese Suchmaschine (bzw. Suchoberfläche) den komplexesten Zugriff auf den Inktomi-Index erlaubt. Die Firma Inktomi bietet keine eigene Suchmaschine an, sondern lizensiert ihre Technologie an Seitenbetreiber. Dabei können diese natürlich auch über die Komplexität der Abfragemöglichkeiten entscheiden.

In den von allen Suchmaschinen angebotenen erweiterten Suchformularen sind in der Regel die gängigsten Einschränkungsmöglichkeiten aufgeführt. Um die volle Funktionalität der jeweiligen Suchmaschine nutzen zu können, ist es jedoch

nötig, die Kommandosprachen zu verwenden. Im weiteren Text werden die einzelnen Funktionen vorgestellt, eine Übersicht, welche auch die Kommandos in den unterschiedlichen Abfragesprachen enthält, gibt die Tabelle am Ende des Beitrags.

Boolesche Operatoren

Bei allen untersuchten Suchmaschinen werden Mehrwortanfragen automatisch mit AND verknüpft. Die Verknüpfung von Suchbegriffen mittels des OR-Operators ist durchweg möglich.
Der NOT-Operator wird in der Regel durch das Minuszeichen vor einem Wort ausgedrückt. Bei *All The Web, Alta Vista, Hotbot* und *Fireball* kann der Operator auch in Worten eingegeben werden, was insbesondere bei der Klammerung innerhalb von Suchargumenten von Bedeutung ist.
Eine volle Unterstützung der Booleschen Operatoren, also auch die Möglichkeit, komplexe Suchargumente mit Hilfe von Klammersetzung zu bilden, bieten *All The Web, Alta Vista, Hotbot* und *Fireball*. Bei *Google* und *Teoma* sind komplexe Suchanfragen nur schwer möglich, lassen sich aber teils durch nicht regelkonforme Syntax simulieren (vgl. Tabelle 1).

Tabelle 1: Search Engine Showdown Analysis: Boolean Searching on Google (Notess 2000)

For this operation	Try this on Google
x AND y	x y
x OR y	x OR y
x AND (y OR z)	x y OR z
(x AND y) OR (z AND q)	not yet possible
(x OR y) AND (z OR q)	x OR y z OR q
x AND (y OR z OR q)	x y OR z OR q
x AND (y OR z) AND q	x y OR z q

Proximity-Operatoren

Die Suche mit dem Abstandsoperator NEAR ist bei *Alta Vista* und *Fireball* möglich. Der voreingestellte Wortabstand beträgt dabei maximal zehn Wörter. Mit dem WITHIN-Operator bei *Alta Vista* lässt sich der maximale Wortabstand auch selbst bestimmen. Die Syntax lautet hier: `{Suchbegriff 1} within {Wert} {Suchbegriff 2}`. Dabei ist auf das Leerzeichen zwischen within und dem Wert zu achten. Die anderen untersuchten Suchmaschinen unterstützen keinerlei Abstandsoperatoren, abgesehen von (wenigstens manchmal brauchbaren) Hilfsmitteln wie dem Platzhalter in der Phrasensuche bei Google (siehe „Phrasensuche").

Phrasensuche

Alle genannten Suchmaschinen unterstützen die Phrasensuche durch das Setzen von Anführungszeichen. Sowohl Google als auch *Alta Vista* erlauben eine Art Trunkierung innerhalb der Phrasensuche: Wird ein Platzhalten (in beiden Fällen: *) gesetzt, so ersetzt dieser genau ein Wort. Nützlich ist diese Funktion vor allem für Suchanfragen, bei denen entweder ein Wort nicht bekannt ist, gleichzeitig nach alter und neuer Rechtschreibung gesucht werden soll oder aber bewusst Schreibfehler mit in die Anfrage eingeschlossen werden sollen.
Es besteht auch die Möglichkeit, den Platzhalter – getrennt durch Leerzeichen – mehrmals hintereinander zu setzen, wobei jeder Stern für genau ein Wort steht.

Synonyme

Allein bei *Google* besteht die Möglichkeit, eine Suchanfrage um Synonyme zu erweitern (vgl. Lewandowski 2003). Dazu wird dem jeweiligen Suchbegriff das Tilde-Zeichen (~) vorangestellt. Die Synonyme werden bei *Google* automatisch generiert. Man sollte sich also trotz der teilweise nützlichen Ergebnisse nicht allein auf diese Funktion verlassen, sondern sie eher als Anregung für die eigenständige Suche nach weiteren Synonymen betrachten.

Trunkierung

Die einzigen Suchmaschinen, die eine Trunkierung anbieten, sind Alta Vista und Fireball. Mit dem Sternchen werden beliebig viele Zeichen am Ende des Worts ersetzt. Es müssen allerdings mindestens drei Zeichen vor dem Trunkierungszeichen stehen. Bei Alta Vista ist ebenso eine Binnentrunkierung möglich. Es wäre zu wünschen, dass auch andere Suchmaschinen eine Trunkierungsfunktion einführen, da ihr Wert unbestritten ist und sie zur Standardausstattung professionelle Information-Retrieval-Systeme zählt.

Feldbeschränkungen

Oft lohnt es sich, die Suche auf bestimmte Felder einzuschränken. Insbesondere eine Suche im Titel des Dokuments ist oft Erfolg versprechend. Zu beachten ist allerdings, dass bei Webseiten die vergebenen Titel oft nicht aussagekräftig sind oder schlicht vergessen wurde, den Titel in das dafür vorgesehene Feld einzutragen. Einige Web-Content-Management-Systeme setzen den Titel auch automatisch, und zwar für alle Seiten einer Website gleich.
Alle untersuchten Suchmaschinen unterstützen die Einschränkung der Suche auf den Titel des Dokuments, entweder über eine Funktion in der erweiterten Suche oder in der Abfragesprache über den Befehl *title*:
Weitere Einschränkungsmöglichkeiten sind die Beschränkung auf die URL der Seite oder einen enthaltenen Linktext. Als einzige Suchmaschine bietet *Fireball* die Möglichkeit, die Suche auf den Inhalt der Metatags Schlagwörter, Autor und Herausgeber einzuschränken.
In manchen Fällen kann auch die Einschränkung der Suchanfrage auf den eigentlichen Text der Seite sinnvoll sein, wobei der in den seitenbeschreibenden Elementen vorkommende Text ausgeschlossen wird. Diese Möglichkeit bieten *Google, All The Web, Alta Vista* und *Fireball*.

Einschränkung nach der Herkunft der Seiten

Einschränkung nach Sprache

Alle hier vorgestellten Suchmaschinen unterstützen die Einschränkung auf eine bestimmte Sprache. Die Dokumente werden im Indexierungsprozess automatisch einer Sprache zugeordnet; dabei sind die Verfahren unterschiedlich, so dass die Zuordnung unterschiedlich zuverlässig ist.
Nützlich ist die Spracheinschränkung insbesondere bei der Suche nach Begriffen, die in mehreren Sprachen gebräuchlich sind sowie bei Akronymen. Sinnvoll kann auch die Einschränkung auf mehrere Sprachen sein, in der Regel auf alle die Sprachen, die man selbst versteht. Diese Möglichkeit bieten *Alta Vista* und *All The Web*. Bei beiden ist die Voreinstellung Englisch und Deutsch; weitere Sprachen können beliebig hinzugefügt werden.

Einschränkung auf eine Top-Level-Domain oder einen Server

Oft ist es notwenig, die Suche auf Dokumente einzuschränken, die auf einem bestimmten Server liegen, aus einem bestimmten Land stammen oder (bei US-amerikanischen Domains) einem bestimmten Bereich wie etwa der Regierung oder dem Hochschulbereich zugehörig sind. So lässt sich eine Suche nach Seiten von amerikanischen Regierungsseiten leicht durch die Beschränkung auf die Top-Level-Domain .gov einschränken. Eine vergleichbare Möglichkeit für deutsche Seiten gibt es nicht.
Alle genannten Suchmaschinen bieten die Möglichkeit, die Suche auf eine Domain oder einen Server zu beschränken. Die Auswahl eines bestimmten Servers ist beispielsweise dann sinnvoll, wenn die Seiten einer Fakultät einer Universität durchsucht werden, dabei aber alle anderen Fakultäten derselben Universität ausgeschlossen werden sollen. Die Einschränkung könnte also in diesem Fall lauten: *site:phil-fak.uni-duesseldorf.de*.
Alle Suchmaschinen außer *Teoma* bieten auch die Möglichkeit, gezielt Seiten eines bestimmten Servers auszuschließen. Diese Funktion wird allerdings nur noch selten benötigt, weil nahezu alle Suchmaschinen inzwischen die Ergebnisse eines Servers clustern, d.h. in der Trefferliste nur zwei Ergebnisse des gleichen Servers anzeigen, allerdings die Möglichkeit bieten, sich durch Anklicken eines Links auch die restlichen Ergebnisse dieses Servers anzeigen zu lassen.
All The Web, Alta Vista und *Fireball* bieten auch die Möglichkeit, mehrere Domains

oder Server in einer Anfrage durch OR zu verbinden. So lassen sich beispielsweise mit *wlan AND (host:heise.de OR host:golem.de)* alle bei den Newsdiensten Heise und Golem veröffentlichten Nachrichten zum Thema WLAN finden. Eine solche Einschränkung ist bei *Google*, *Teoma* und *Hotbot* nicht möglich.
Eine weitere Besonderheit bietet *All The Web*: Hier lässt sich die Suchanfrage auch auf eine bestimmte IP-Adresse oder einen Adressraum einschränken.

Datumsbeschränkung

Die Erfassung des tatsächlichen Datums eines Dokuments durch Suchmaschinen ist als ausgesprochen unzuverlässig anzusehen. Datumsangaben in den Metainformationen eines Dokuments können nicht nur leicht manipuliert werden, sondern ändern sich in vielen Fällen bei der Generierung der Dokumente aus Content-Management-Systemen oder beim Neu-Aufspielen auf den Server. Dennoch sollten die Möglichkeiten, die Treffermenge mittels der Datumseingabe wesentlich zu beschränken, nicht unterschätzt werden.
Prinzipiell bieten alle Suchmaschinen die Möglichkeit der Datumsbeschränkung. Dabei ist zu unterscheiden zwischen der Beschränkung auf aktuelle Dokumente („Dokumente der letzen vier Wochen", „des letzten Jahres", usw.) und einer genauen Bestimmung des Datums.
Google bietet nur die erste Möglichkeit. Der Zeitraum lässt sich auf die letzten drei, sechs oder zwölf Monate einschränken. Tara *Calishain* beschreibt ein Verfahren, auch mit *Google* nach genauen Zeiträumen zu suchen (Calishain/Dornfest 2003, 37), dieses ist allerdings sehr umständlich und für den regulären Gebrauch damit nur schlecht geeignet.
Auch Alta Vista und Fireball bieten vorgegebene Zeiträume an, dazu allerdings auch die Möglichkeit der genauen Angabe des gewünschten Zeitraums. Bei allen Suchmaschinen außer *Google* lassen sich Dokumente finden, die entweder vor, nach oder während eines bestimmten Zeitraums erstellt bzw. aktualisiert wurden.

Dokumenttypen

Dateiformate

Eine Einschränkung auf PDF- oder Microsoft-Office-Dokumente lässt sich bei *Google*, *All The Web* und *Hotbot* vornehmen. Will man nur Postcript-Dateien oder RTF-Dokumente angezeigt bekommen, helfen *Google* oder *All The Web*. Nur *All The Web* bietet eine Einschränkung auf StarOffice, Word Perfect und Flash-Dateien. Der Befehl für die Einschränkung auf einen Dateityp ist in der Regel filetype:, bei *Hotbot* muss das Dateiformat über Ankreuzfelder im erweiterten Suchformular gewählt werden. Dafür können hier (wie auch bei *All The Web*) mehrere Dateiformate ausgewählt werden. Dies ist bei *Google* nicht möglich. *Teoma* und *Fireball* unterstützen generell keine Einschränkungen nach Dateiformaten.

Eingebettete Inhalte

Unter eingebetteten Inhalten werden solche Inhalte verstanden, auf die von einer HTML-Seite aus verwiesen wird, die selbst jedoch keine HTML-Dokumente sind. So handelt es sich in diesen Fällen beispielsweise um zum Text gehörende Video- oder Audiodateien. Diese Suchfunktion wird sicher nicht oft benötigt werden; allerdings lässt sich damit beispielsweise schnell und unkompliziert ein Tonbeispiel für den Gesang des Rotkehlchens finden. Diese Suche wäre ohne die Einschränkung äußerst mühevoll, weil eine große Anzahl Seiten existiert, die sich zwar mit dem Gesang des Rotkehlchens beschäftigen, aber keine Tonbeispiele liefern.
Die Möglichkeit, die Suche auf Seiten, die eingebetteten Inhalt enthalten, einzuschränken, bietet *All The Web*. *Alta Vista* bietet diese Möglichkeit nur für Java-Applets.

Dateigröße

Als einzige Suchmaschine unterstützt *All The Web* eine Suche nach Dokumenten einer bestimmten Länge. Die Dokumentlänge wird hier in Kilobyte bestimmt und die Einschränkung ist möglich nach Dokumenten, die größer oder kleiner als eine bestimmte KB-Zahl sind; außerdem lässt sich die exakte Länge eines Dokuments angeben. Dabei sagt die KB-Zahl natürlich nur eingeschränkt etwas über die tatsächliche Länge des Dokuments aus, weil auf der Seite vorhandene Navigationselemente, etc. die KB-Zahl erhöhen. Bei der Recherche ist generell zu beachten, dass Google Dokumente nur bis zu einer Länge von 100 KB indexiert. Begriffe, die erst nach dieser Grenze im Dokument vorkommen, sind über Google nicht auffindbar.

Geographische Einschränkungen

Alle Suchmaschinen bieten die Einschränkung auf ein bestimmtes Land durch die Beschränkung auf eine bestimmte Top-Level-Domain. Allerdings werden hier Seiten ausgeschlossen, die zwar in einem bestimmten Land erstellt wurden, beispielsweise aber auf einer .com-Domain liegen. Hier ist die Einschränkung nach der Sprache sinnvoller.
Ist man auf der Suche nach Dokumenten, die einer größeren geographischen Einheit zugehörig sind, so sollte man *Teoma, All The Web* oder *Hotbot* nutzen: Hier werden unterschiedliche Regionen (z.B. Europa, Asien) zur Einschränkung angeboten. Sinnvoll ist dies beispielsweise, wenn man sich über neuere Entwicklungen im Bereich der Mobilkommunikation in Asien informieren möchte.

Verlinkungskontrolle

Um zu überprüfen, wer auf die eigene Website/die eigenen Seiten linkt, ist es bei den meisten Suchmaschinen möglich, eine solche Suche auszuführen. Gerade in der momentanen Situation, in der Linkanalyse eines der bedeutendsten Rankingkriterien für Suchmaschinen ist, sind solche Recherchen von enormer Bedeutung. Auf der Basis der Ergebnisse können Site-Betreiber gebeten werden, ihre Linktexte gemäß den eigenen Wünschen anzupassen und solche Anbieter identifiziert werden, die von einem thematischen Standpunkt her einen Link setzen könnten oder sollten, dies aber bisher noch nicht getan haben.
Die Syntax lautet in der Regel *link:* und wird von allen behandelten Suchmaschinen außer *Teoma* und *Hotbot* unterstützt.

Kombinierbarkeit der Einschränkungen

Anders als bei den Retrievalsystemen kommerzieller Datenbanken lassen sich bei Suchmaschinen oftmals die Operatoren nicht beliebig kombinieren.

Nur *Alta Vista, All The Web* und *Fireball* erlauben die beliebige Kombination von Operatoren und Feldbeschränkungen. Da solche Kombinationsmöglichkeiten eine unabdingbare Voraussetzung für die professionelle Recherche sind, ist festzustellen, dass sich ausschließlich diese Suchmaschinen – trotz anderweitig bestehender Defizite – für komplexe Recherchen eignen.

Verbesserung von Suchanfragen

Rechtschreibkontrolle

Ein wichtiges Tool für die Recherche ist die automatische Rechtschreibkontrolle, die von *Google, All The Web* und *Teoma* angeboten wird. Leicht vertippt man sich einmal, insbesondere bei komplexen Suchargumenten. Vor allem wenn einzelne Wörter mit OR verknüpft werden, werden Schreibfehler oft nicht bemerkt, verkleinern aber die Ergebnismenge.
Die Korrekturvorschläge, die die Suchmaschinen anbieten, werden automatisch generiert, was einerseits den Vorteil hat, dass auch Wörter, die nicht in einem Wörterbuch hinterlegt sind, verbessert werden können. Dafür kommt es aber manchmal auch zu unsinnigen Vorschlägen (Bspw. Eingabe: filetype:msword, Vorschlag: „meinten Sie: *filetype: sword?*"). In allen Fällen werden die Eingaben des Nutzers aber nicht automatisch verbessert, sondern es werden erst die Suchergebnisse aufgrund der (inkorrekt) gestellten Anfrage ausgegeben. Die Rechtschreibkontrolle liefert also nur Vorschläge.

Vorschläge zur Verbesserung der Suchanfrage

Eine wichtige Methode, Suchanfragen einzuschränken, kann in Retrievaltests nicht berücksichtigt werden: die Einschränkung in der Treffermenge in einem zweiten Schritt, also nach der Anzeige der Trefferliste. Dazu werden von der jeweiligen Suchmaschine automatisch Begriffe gefunden, die auf vielen der gefundenen Webseiten vorkommen und sich eventuell dazu eignen, das Thema weiter einzuschränken. Solche Verfahren werden von *Teoma* und *Alta Vista („Alta Vista Prisma")* verwendet. Oft lässt sich mit ihrer Hilfe in einem iterativen Prozess (auch in Kombination mit den oben beschriebenen Einschränkungsmöglichkeiten) die Anfrage so weit einschränken, dass eine überschaubare Treffermenge übrig bleibt, die dann tatsächlich vollständig gesichtet werden kann.
Ein ähnliches Verfahren ist die Clusteranalyse. Hier werden die Ergebnisse in Klassen gruppiert, die aufgrund gewisser Gemeinsamkeiten der Dokumente gebildet werden. Vorreiter dieses Verfahrens im Suchmaschinen-Bereich war die inzwischen eingestellte Suchmaschine Northern Light, (eingeschränkt) angewendet wird das Verfahren bei All The Web. Allerdings gibt es auch Suchmaschinen, deren Stärken gerade bei diesem Verfahren liegen, die jedoch aufgrund der o.g. Auswahlkriterien im Rahmen dieses Aufsatzes nicht behandelt werden (z.B. Vivisimo und Kartoo).

Tabelle 2: Abfragemöglichkeiten in den Suchmaschinen Teoma, Google, All The Web, Alta Vista, Hotbot und Fireball

	Teoma
URL	www.teoma.com
Operatoren	
AND	ja
OR	ja
AND NOT	ja (mit -)
Proximity-Operatoren	
NEAR	nein
Phrasensuche	ja
Trunkierung	
	-
Kombinations-möglichkeiten	
Vollständige Unterstützung durch Klammerung	nein
Einschränkung auf Text...	
... im Titel der Seite	intitle:
... im Hauptteil der Seite (Fließtext)	-
... in der URL der Seite	Inurl:
... in einem Link auf d. Seite	-
... im Metatag Keyword	-
... im Metatag Author	-
... im Metatag Publisher	
Einschränkung nach Herkunft der Seiten	
Sprachauswahl	Lang: (eine Spra
Domain	Einschließen *Site:*
IP-Adresse	-
Ähnliche Seiten	-
Geographische Einschränkungen	
Länderauswahl	Über Sprache o
Einschränkung auf größere geographische Einheiten	*Geoloc:*
Datumsbeschränkungen	
Periodisch	*last:*
Exakt	Vor / nach / zw *afterdate / befo betweendate*
Einschränkung nach Dokumenttypen	
Dateiformate	-
Eingebettete Inhalte	-
Dateigröße	-

Google www.google.de	All The Web www.alltheweb.com	Alta Vista www.altavista.de	Hotbot (Inktomi) www.hotbot.de	Fireball www.fireball.de
ja	ja	ja	Über Auswahlmenü	ja
eingeschränkt	ja	ja	Über Auswahlmenü	ja
ja (mit -)	Ja: *andnot*	ja	Über Auswahlmenü	ja
nein	nein	ja (Standard-Abstand bei NEAR: 10 Wörter). *Within:* {Wert}: maximaler Abstand zwischen den Wörtern bestimmbar	Nein	ja (Abstand: 10 Wörter)
ja	ja	ja	ja	Ja
-	Nur im Domainnamen mit *	Rechts- und Binnen-trunkierung mit *	-	Rechtstrunkierung mit *
nein	ja	ja	nein	ja
Intitle: (über Auswahlmenü)	Title: (über Auswahlmenü)	Title: Text:	ja (über Menü) -	Title: Text:
Inurl: Link: - -	url: Link: - -	url: Link: - -	- - - -	url: Link: Keyword: Author: Publisher:
einzelne Einschließen, ausschließen	Einzelne; mehrere Language: Einschließen, ausschließen Mehrere einschließen möglich	Einzelne, mehrere Einschließen, ausschließen, mehrere *Domain:* *Host:*	Einzelne, mehrere Einschließen, ausschließen	Einzelne Einschließen, ausschließen, mehrere Host:
- ja	Einschließen -	- *Like:*	- -	- -
Über Sprache oder TLD -	Über Sprache oder TLD ja	Über Sprache oder TLD -	ja; wenige Länder ja	Über Sprache oder TLD -
3 Monate, 6 Monate, 1 Jahr		1 Woche, 2 Wochen, 1 Monat, 4 Monate, 8 Monate, 1 Jahr		1 Tag, 1 Woche, 2 Wochen, 1 Monat
	Vor / nach / zwischen	Vor / nach / zwischen	Vor / nach / zwischen	Vor / nach / zwischen
Filetype: PDF PS DOC XLS PPT RTF	Filetype: PDF Flash MSWORD RTF POWERPOINT EXCEL POSTCRIPT WORDPERFECT STAROFFICE	PDF	HTML PDF TEXT WORD EXCEL PPT	-
-	Einschließen / ausschließen Bilder Audio Video Audio-/Video-Streams Flash Java Javascript VB Script	Java Applet: Object:	-	-
-	Größer / kleiner / gleich Filesize:>{Wert} Filesize:[{Wert1},{Wert2}]	-	-	-

Eine weitere Besonderheit bietet Teoma: Diese Suchmaschine ist aufgrund eines besonderen Algorithmus in der Lage, hochwertige zur Suchanfrage passende Linksammlungen zu identifizieren („link collections from experts and enthusiasts"). Durch diese Funktion wird man oft auf die wichtigsten Quellen in einem Themenfeld gelenkt, ohne als Nutzer einen großen Aufwand betrieben zu haben.

Schlussbemerkung

Die Qualität von Suchmaschinen lässt sich nicht allein durch die Überprüfung von Ein- und Mehrwortanfragen messen. Aus diesem Grund ist bei Retrievaltests stets zu fragen, ob die aus ihnen gewonnenen Aussagen auch dann Gültigkeit besitzen, wenn komplexe Suchanfragen betrachtet werden. Die Antwort lautet, dass diejenige Suchmaschine, die in Retrievaltests als „die Beste" identifiziert wird, für manche Suchanfragen schlicht ungeeignet sein kann oder aber, dass bestimmte Suchfragen mit einer anderen Suchmaschine schneller bzw. effizienter gelöst werden können.

Literatur

Calishain, Tara; *Dornfest*, Rael: Google Hacks: 100 Industrial-Stregth Tips & Tools. Sebastopol [u.a.], 2003

Clay, Bruce (2003): Search Engine Relationship Chart. http://www.bruceclay.com/searchenginechart.pdf [14.11.2003]

Griesbaum, Joachim; *Rittberger*, Marc; *Bekavac*; Bernhard (2002): Deutsche Suchmaschinen im Vergleich: AltaVista.de, Fireball.de, Google.de und Lycos.de. In: Information und Mobilität. Optimierung und Vermeidung von Mobilität durch Information. Proceedings des 8. Internationalen Sysmposiums für Informationswissenschaft. *Hammwöhner*, R., *Wolff*, C., und *Womser-Hacker*, C. (Hrsg.), 201-223

Hock, Ran: Web Search Engines: (More) features & commands. Online 24(2000)3. http://www.findarticles.com/cf_0/m1388/3_24/61640524/print.jhtml [14.11.2003]

Karzaunikat, Stefan (2003): Google zugemüllt: Spam überschwemmt die Suchergebnisse. In: c't (2003)20, 88-92

Lewandowski, Dirk (2003): Bessere Suchmöglichkeiten durch neuen Operator? In: Password (2003)9, 36

Notess, Greg (2003): Unusual Power Web Searching Commands. In: Online 27(2003)6, 40-42

Notess, Greg (2000): Search Engine Showdown Analysis: Boolean Searching on Google. http://www.searchengineshowdown.com/features/google/google-boolean.html [14.11.2003]

Silverstein, Craig; *Henzinger*, Monika; *Marais*, Hannes; *Moricz*, Michael: Analysis of a Very Large Alta Vista Query Log. Digital Systems Research Center Technical Note 1998-014

Suchfibel (2003): Das Beziehungsgefecht der Suchmaschinen. http://www.suchfibel.de/5technik/ suchmaschinen_beziehungen.htm [14.11.2003]

Sullivan, Danny (2003): Search Engine Sizes. http://www .sewatch.com/reports/article.php/2156481 [14.11.2003]

Veritest (2003): Inktomi Corp.: Web Search Relevance Test. http://www.veritest.com/clients/reports/inktomi/ inktomi_web_search_test.pdf [14.11.2003]

Suchmaschine, Information Retrieval, Recall, Qualität, Test, Bewertung

Der Autor

Dirk Lewandowski

ist bei der NRW Medien GmbH in Düsseldorf als Researcher tätig sowie als Lehrbeauftragter an der Universität Düsseldorf und der FH Köln. Er betreut die Rubrik „Suchmaschinen-News" in Password und forscht im Rahmen seines Dissertationsvorhabens an Verbesserungsmöglichkeiten für algorithmische Suchmaschinen.

Merkurstrasse 66, 40223 Düsseldorf
E-Mail: dirk.lewandowski@uni-duesseldorf.de