

Suchmaschinen-Update

Markttrends und Entwicklungsperspektiven bei WWW-Universalsuchmaschinen

Dirk Lewandowski
NRW Medien GmbH, Düsseldorf

Der Markt für WWW-Universalsuchdienste hat seit dem Jahr 2002 bedeutende Änderungen erfahren. Einige Suchmaschinen haben ihren Betrieb eingestellt, insgesamt kam es zu einer Konsolidierung des Marktes. Nicht zuletzt durch Aufkäufe bisher unabhängiger Betreiber fand eine weitere Konzentration statt.

Seitens der Benutzer besteht weitgehend das Bild, mit Hilfe der dominierenden Suchmaschine *Google* sämtliche Informationen des WWW auf einfache Weise finden zu können. Es werden Beispiele für Suchanfragen vorgestellt, die sich ausschließlich oder aber besser mit Hilfe anderer Suchdienste beantworten lassen. Ein Ausblick in die Zukunft identifiziert die intuitive Benutzerführung, die Einbindung von *Invisible-Web-Content* und den Umgang mit *paid content* als die wesentlichen Herausforderungen für die näheren Zukunft.

Das Thema Websuche erfreut sich weiterhin großer Popularität. Ziel dieses Aufsatzes ist, einerseits die Strukturen des Marktes für allgemeine Suchmaschinen darzustellen, andererseits die technischen Fortschritte dieser Suchmaschinen aufzuzeigen. Dazu werden auch Funktionen, die nur von einzelnen Suchwerkzeugen angeboten werden, vorgestellt.

Außen vor bleiben spezialisierte Suchdienste, die sich entweder auf bestimmte Themenfelder, Regionen oder bestimmte Inhalte bzw. Formate wie Audio oder Video beschränken (vgl. dazu Gelernter 2003). In Bezug auf den durch Suchanfragen generierten Traffic besteht eine klare Dominanz der Universalsuchdienste. Es ist weiter davon auszugehen, daß der Nutzer es vorzieht, nur einen Suchdienst für alle seine Recherchen zu verwenden. Insofern besteht bei den Suchmaschinen eine Portal-Tendenz, allerdings nicht wie noch vor wenigen Jahren durch die Ergänzung um suchfremde Dienste, sondern durch die Erweiterung des Suchhorizonts bzw. der Erweiterung um weitere Dokumentkollektionen, die durchaus von spezialisierten Anbietern stammen können (Directory, sponsored Listings, Multimedia, News).

Fortlaufende Entwicklung

Auch im letzten Jahr haben sich die sich schon in den letzten Jahren abzeichnenden Trends fortgesetzt¹: die Indexgrößen der wesentlichen Suchmaschinen sind weiter gewachsen (die großen Indizes liegen bei etwa zwei bis drei Milliarden Dokumenten) und die Aktualität der Datenbestände konnte weiter verbessert werden (Notess 2002). Auch die Entwicklung hin zur Indexierung von immer mehr Dateitypen setzte sich fort. PDF und die Formate der Microsoft-Office-Produkte dürfen inzwischen als Standard gelten, Multimedia-Formate wie Macromedia Flash werden bisher einzig von *FAST / AlltheWeb* erfaßt.

Auch in Hinblick auf das Ranking der Trefferlisten wurden weitere Fortschritte erzielt. Während *Google* längere Zeit als nahezu konkurrenzlos galt, hat *AltaVista*, aber vor allem *AlltheWeb* im letzten Jahr das Ranking wesentlich verbessert. Allerdings erfreut sich keine dieser beiden Suchmaschinen einer ähnlichen Popularität wie der klare Marktführer *Google*. *AlltheWeb* ist als eigenständige Suchmaschine außerhalb der Community der Information Professionals sogar weitgehend unbekannt.

¹ zu den Entwicklungen der Jahre 2001 und 2002 siehe Lewandowski (2002)

Trotz der aufgezeigten positiven Entwicklungen dürfte die Abdeckung der Webinhalte durch die Suchmaschinen weiter abnehmen. Einerseits wächst die Anzahl konventioneller Webseiten (HTML, Standardformate) schneller als die Indizes der Suchmaschinen, andererseits breitet sich das sog. „Invisible Web“ mit extremer Geschwindigkeit aus. Zwar mögen die Schätzungen über dessen Größe im Allgemeinen zu hoch angesetzt sein², trotzdem ist dieser Bereich quantitativ und vor allem qualitativ³ von hoher Bedeutung.

Das Anwachsen der Indizes darf als wesentliche Hürde für den Markteinstieg neuer Suchmaschinen gelten. Um gegenüber den etablierten Anbietern konkurrenzfähig zu sein, bedarf es eines umfangreichen Datenbestandes, dessen Aufbau mit enormem Hardware- und Geldbedarf verbunden ist.

Für die Beurteilung der Indexgrößen ist man, von wenigen Untersuchungen abgesehen, weitgehend auf die Angaben der Betreiber angewiesen. Da jedoch die Zahl der indexierten Dokumente als für den Laien einzig objektives Vergleichskriterium gelten muß, ist in Bezug auf deren Glaubwürdigkeit Vorsicht geboten.

Eine von Greg Notess regelmäßig durchgeführte Untersuchung der Indexgrößen (Notess 2003) zeigt, daß die von *AlltheWeb* und *Google* angegebenen Werte stimmen dürften, während der *Inktomi*-Index in Wahrheit deutlich kleiner sein dürfte als die angegebenen drei Milliarden Dokumente. Bei *AltaVista* ergeben sich in der Untersuchung Werte, die deutlich über der von *AltaVista* angegebenen Marke von einer Milliarde Dokumenten liegen.

In der Diskussion um die Indexgrößen und ihre Aussagekraft in Bezug auf die Qualität einzelner Suchmaschinen wird meist das Grundproblem der großen, schwach strukturierten Dokumentenmengen und ihres Rankings aus den Augen verloren. Fraglich ist nämlich, ob Rankingverfahren überhaupt konsequent zu qualitativ hochwertigen Ergebnissen führen können, wenn Trefferlisten ausgewertet werden, die einige tausend bis Millionen Dokumente enthalten.⁴

² Bergman (2001) schätzt die Zahl der Dokumente im *surface web* im März 2000 auf etwa eine Milliarde, die des *deep web* (= *invisible web*) hingegen auf etwa 550 Milliarden individuelle Dokumente. Dabei ist allerdings zu bedenken, daß sich die Inhalte von Datenbanken nur schwer in eine mit den Dokumenten des *surface web* vergleichbare Zahl bringen lassen, weil es sich hierbei vornehmlich um Datenbankinhalte handelt. Jeden Datensatz als individuelles Dokument zu zählen, erscheint wenig sinnvoll. Dazu kommt die Problematik der Dubletten: so mag allein der Karlsruher Virtuelle Katalog Zugriff auf 75 Millionen bibliographische Datensätze bieten, die Zahl der überschneidungsfreien Datensatzmenge dürfte jedoch weit geringer sein.

³ Dabei wird angenommen, daß die Inhalte des Invisible Web (also vor allem Datenbank-Inhalte) tendenziell eine höhere Qualität haben als die Inhalte von HTML-Seiten. Dies hängt mit den Hürden für die Erstellung einer Datenbank, der Notwendigkeit einer Datenbank erst ab einer signifikanten Dokumentenmenge und der tendenziell höheren Wahrscheinlichkeit, daß die Inhalte von professionellen Anbietern (seien diese gewinnorientiert oder non-profit) produziert werden, zusammen.

⁴ so zeigt eine Untersuchung deutschsprachiger Suchmaschinen, daß die Mean Average Precision in den Top 20 der Trefferlisten bei den Suchmaschinen *AltaVista*, *Fireball*, *Lycos* und *Google* zwischen 0,396 und 0,551 liegt. D.h. nur etwa jeder zweite Treffer der Top 20 in den Trefferlisten des Testsiegers *Google* wird als relevant eingestuft bzw. verweist auf ein relevantes Dokument. (Griesbaum et al. 2002)

Marktentwicklung

Während noch 2001 relativ stabile Marktanteile bestanden, die nur eine gewisse Tendenz zur Verstärkung des Anteils von *Google* zeigten, hat *Google* im Lauf des Jahres 2002 eine enorme Marktmacht errungen. Die Suchmaschinennutzer verlassen sich zunehmend auf *Google* als einzige Quelle ihrer Recherche, was dazu führt, daß Informationen, die im *Google*-Index nicht vorhanden sind, vielfach nicht wahrgenommen werden. Dazu kommt, daß die erweiterten Suchmöglichkeiten von *Google* nur sehr begrenzt sind (keine vollständige Unterstützung der Boole'schen Operatoren, nur eingeschränkte zeitliche Suche, keine Einschränkung nach *embedded content*, usw.). Da *Google* auch aus den meisten Suchmaschinen-Tests als Sieger hervorgeht (vgl. z.B. Griesbaum et al. 2002, Stiftung Warentest 2003), besteht vielfach die Meinung, daß *Google* für sämtliche Suchanfragen die geeignetste Lösung darstellt.

Diese Entwicklungen zusammen mit dem Verkauf der *Google*-Ergebnisse an bedeutende Portalseiten (wie z.B. *Yahoo*) haben dafür gesorgt, daß – gemessen an der Menge der Suchanfragen – ein Quasi-Monopol im Bereich Web-Suche entstanden ist. Daraus ergeben sich Probleme wie diejenigen kleiner Unternehmen, die für ihr Fortbestehen auf das positive Ranking durch *Google* angewiesen sind. Dies führte bis zu Prozessen um den *PageRank*-Wert und der Frage nach der Verantwortung eines (angeblichen) Monopolisten.⁵

Im Zuge einer Marktbereinigung haben einige renommierte Suchmaschinen aufgegeben, andere werden ohne eigenen Index weitergeführt. Generell läßt sich die Tendenz feststellen, daß immer mehr Suchmaschinen ihren Index zukaufen.⁶ Den Weltmarkt führen inzwischen vier Anbieter umfangreicher und qualitativ hochwertiger Indizes an. Dies sind neben *Google* die Indizes von *FAST* (u.a. eingesetzt von *Alltheweb*, *Lycos*), *Inktomi* (*MSN Search*, *HotBot*) und *Alta Vista*. Weitere Datenbestände, die durch ihre Größe durchaus relevant sein dürften (wie z.B. *Teoma*), spielen bei der Vermarktung von Suchergebnissen nur eine untergeordnete Rolle.

Abbildung 1 zeigt eine Momentaufnahme der Beziehungen der wichtigsten Suchseiten untereinander. Einerseits zeigt sich, daß keines der großen Portale mehr einen eigenen Datenbestand für seine Suchanfragen verwendet.⁷ Andererseits zeigt sich, daß auch diejenigen Suchmaschinen, die eigene Datenbestände unterhalten, weitere Bestände wie Ergebnisse aus Verzeichnissen und Werbung von anderen Anbietern zukaufen.

Im Gegensatz zu ehemals eigenständigen Suchmaschinen, die inzwischen Ergebnisse *einer* anderen Suchmaschine liefern oder zu einer Metasuchmaschine geworden sind, ist *HotBot* einen anderen, bisher von keiner (ehemals) großen Suchmaschine beschrittenen Weg gegangen.

Die Suchmaschine *HotBot*, eine Zeitlang ausgesprochen beliebt wegen ihrer erweiterten Suchfunktionen, hatte in den letzten Jahren einen Verfall erfahren, nachdem sie unter das Dach von *Lycos* gekommen war. *HotBot* verwendete bis 2002 den Index von *Inktomi*, seit dem Relaunch kann bei der Suche mittels eines Radiobuttons zwischen den Datenbeständen von *Inktomi*, *Google*, *FAST* und *Teoma*

⁵ Es sei hier auch auf die Diskussion hingewiesen, inwieweit *Google* Nutzungsdaten akkumuliert und IP-bezogen auswertet.

⁶ Suchdienste, die früher eine Suchmaschine mit eigener Datenbank betrieben haben, die Suchergebnisse aber mittlerweile zukaufen, sind z.B. *Lycos* und *Excite*.

⁷ Das Gleiche gilt für die deutschsprachigen Portale. Beispielsweise beziehen *T-Online* und *Lycos* ihre primären Suchergebnisse von *FAST*, *Web.de* von *Google*.

gewählt werden.⁸ Es handelt sich dabei aber nicht um eine weitere Metasuchmaschine, sondern um ein Suchinterface, mit dem sich leicht zwischen den Ergebnissen einzelner Suchmaschinen hin- und herwechseln läßt.⁹

Die Finanzierung der Suchmaschinen erfolgt nicht mehr wie noch in den vergangenen Jahren primäre über Bannerwerbung, sondern über gezielt an die Suchanfrage angepaßte Textanzeigen. Dabei ist es üblich, daß die Werbetreibenden nicht für die Einblendung der Werbung selbst, sondern erst für deren Anklicken bezahlen (*pay-per-click advertising*).

Eine Einnahmequelle für die Suchmaschinenbetreiber ist aber nicht nur die Plazierung der Werbung auf ihren eigenen Ergebnisseiten, sondern auch deren Vermarktung auf Portalseiten und sonstigen Suchangeboten.¹⁰ Die beiden weltweit führenden Anbieter für die Vermarktung von Textanzeigen sind *Google* und *Overture*, von Bedeutung auf dem europäischen Markt ist weiterhin *Espotting*.

Während *Google* schon in der Vergangenheit seinen Partnerseiten eine „All-in-One“-Lösung anbieten konnte, die sowohl die tatsächlichen Ergebnisse der Suche als auch die Werbeplätze beinhaltet, und diese auch erfolgreich an Portale wie *Yahoo* vermarkten konnte, konnte *Overture* bisher einzig Textanzeigen anbieten und nicht über die Plazierung der Anzeigen im Suchumfeld entscheiden. Eine eigene Suchtechnologie hatte die Firma nicht entwickelt. Um gegenüber *Google* konkurrenzfähig zu bleiben, kaufte die Firma im Februar 2003 sowohl *AltaVista* als auch den Bereich Websearch von *FAST* (vgl. Sullivan 2003). Interessant ist diese Entwicklung weniger wegen der Frage nach den Marktanteilen der Textanzeigenvermarkter als wegen der Frage nach der Qualität und Neutralität der angebotenen Trefferlisten. Ein Grundannahme bei der Benutzung von Suchmaschinen ist, daß diese durch die vollautomatisierte Erfassung und Erschließung der Dokumente und die durch Algorithmen gesteuerte Aufbereitung der Suchergebnisse unbeeinflusste Resultate liefern. Dies ist zugleich die Grundbedingung für das notwendige Vertrauen durch die Nutzer.¹¹

Während die Textanzeigen bei *Google* und den entsprechenden Partnerseiten relativ klar von den Suchergebnissen abgegrenzt sind, fand schon in der Vergangenheit bei *Overture* und seinen Partnern eine Vermischung von Trefferlisten und bezahltem Werbeplatz statt. Die bezahlten Treffer, die vor den eigentlichen Ergebnislisten gezeigt werden, sind oft nur unklar gekennzeichnet (bspw. „Overture Resultat“ bei *T-Online*).

Abzuwarten bleibt, wie *Overture* seine Textanzeigen auf den Seiten von *Alltheweb* und *AltaVista* plazieren wird. Es ist denkbar, daß insbesondere auf der relativ gut

⁸ Diese Veränderungen betreffen die Seite www.hotbot.com. Die deutschsprachige Version greift weiterhin ausschließlich auf den Inktomi-Index zu.

⁹ Die Veränderungen wurden durchaus positiv aufgenommen, so etwa von Bates (2003, 80): „The new HotBot is a great tool. [...] When I feel that I need to cover the waterfront and find as many relevant sites as possible on a topic, HotBot is my new first choice.“

¹⁰ Seit März 2003 bietet *Google* auch die Vermarktung von inhaltsbezogenen Textanzeigen auf suchfremden Angeboten an (u.a. bei Online-Zeitschriften).

¹¹ so Kuhlen (1999, 125): „Nicht nur werden in der (Informations)Moderne die [...] Situationen so komplex, daß wir uns in ihnen nicht aus eigener Kraft informationell sicher bewegen können, auch die Welt der Informationsmaschinen und informationellen Assistenz wird so komplex, daß wir uns ihrer Leistung kaum ausreichend vergewissern können, weder hinsichtlich dessen, was überhaupt auf den Informationsmärkten verfügbar ist, noch bezüglich der Einschätzung des Wahrheitswertes und der Relevanz dessen, was uns andere an Informationen liefern.“

frequentierten *AltaVista*-Seite eine zunehmende Vermischung zwischen Suchergebnissen und Werbung stattfinden wird. Diese Entwicklung ist an sich nicht neu, gewinnt jedoch an Bedeutung, da nun erstmals Suchmaschinen, die bisher ein hohes Ansehen bezüglich ihrer Qualität und Seriosität genossen, von einem Werbevermarkter übernommen wurden. Für weitere Besorgnis sorgte die Ankündigung von *Overture*, zukünftig eine Partnerschaft mit dem hochumstrittenen *Gator Advertising and Information Network* eingehen zu wollen. Die Software von *Gator*, die in vielen kostenlosen Programmen (wie z.B. beliebten Filesharing-Tools) enthalten ist, spioniert das Nutzerverhalten aus und blendet diesem Pop-Up-Werbung auf dem Bildschirm des Nutzers ein.

Ein weiterer Aufkauf eines bedeutenden Suchmaschinen-Indizes ist im Zuge dieser Entwicklungen schon fast wieder vergessen. Im Dezember 2002 hatte *Yahoo* die Firma *Inktomi* gekauft und verfügt damit nun auch über eine eigene crawler-basierte Suchmaschine. Bisher bezieht *Yahoo* seine primären Ergebnisse von *Google*. Es ist davon auszugehen, daß in Zukunft statt dessen eine weiterentwickelte *Inktomi*-Lösung auf den *Yahoo*-Seiten verwendet wird.

Die genannten Aufkäufe zeigen eine Entwicklung hin zu integrierten Lösungen für Portalseiten: zukünftig wird es einigen wenigen Firmen möglich sein, „All-in-One“-Lösungen bestehend aus Suchergebnissen und Werbeplatzierungen anzubieten.

Neben den international operierenden Suchmaschinen hatten in der Vergangenheit deutschsprachige Suchmaschinen eine besondere Bedeutung, da sie eigene, speziell auf den deutschen Sprachraum zugeschnittene Datenbestände aufbauten, die oftmals mehr deutschsprachige Dokumente enthielten als die Indizes der internationalen Anbieter. Im Zuge des rasanten Wachstums der internationalen Suchmaschinen besteht dieser Vorteil inzwischen nicht mehr (vgl. Griesbaum et al. 2002).¹² Suchmaschinenentwicklungen aus dem deutschsprachigen Raum haben inzwischen auch für diesen abgegrenzten Sprachraum keine Bedeutung mehr.

Spezielle Suchfunktionen einzelner Suchmaschinen

Auch wenn zum gegenwärtigen Zeitpunkt nahezu alle Untersuchungen belegen, daß sich mit *Google* die besten Ergebnisse erzielen lassen, so gilt dies jedoch im wesentlichen für relativ einfach strukturierte Suchanfragen, wie sie der mit Suchtechniken wenig vertraute Nutzer stellt.

Allerdings sind viele komplexe Suchanfragen mit *Google* schlicht nicht möglich bzw. lassen sich einfacher mit anderen Suchwerkzeugen durchführen. Diese gilt einerseits für die Arbeit mit Boole'schen Operatoren, welche nur von *AlltheWeb* und *AltaVista* unterstützt werden, Abstandsoperatoren (ebenfalls bei *Alltheweb* und *AltaVista*), Trunkierung (nur bei *AltaVista*) und Feldsuchen (nur eingeschränkt bei *Google*, mehr Möglichkeiten bei *Alltheweb* und *AltaVista*). Generell läßt sich feststellen, daß *AlltheWeb* und *AltaVista* über die mächtigsten Abfragesprachen im Bereich der Websuche verfügen.¹³ Insofern sind sie gerade für Information Professionals besonders interessant.

¹² Auch die explizit deutschsprachigen Suchmaschinen verwenden inzwischen weitgehend Indizes internationaler Fremdanbieter.

¹³ Einen Überblick über die Suchsprachen findet sich unter http://www.alltheweb.com/help/faqs/query_language.html [18.3.2003] bzw. http://de.Altavista.com/help/adv_search/syntax [18.3.2003].

Weitere nützliche Funktionen lassen sich auf den Seiten der erweiterten Suche bei AlltheWeb finden:

- regionale Einschränkung: die Suche kann nicht nur nach der Sprache oder über die Domainendung nach dem Land eingegrenzt werden, sondern auch nach größeren geographischen Einheiten (z.B. Europa, mittlerer Osten).
- genaue Datumsbeschränkung: sowohl bei *AlltheWeb* als auch bei *AltaVista* läßt sich das Datum der zu suchenden Seiten genau einschränken, während bei *Google* nur eine sehr grobe Einschränkung möglich ist.
- *embedded content*: *AlltheWeb* bietet die Möglichkeit, die Suche auf Seiten zu beschränken, in denen ein gewisser Dateityp eingebettet ist, oder aber solche, die einen bestimmten Dateityp nicht enthalten. Mit dieser Einschränkung läßt sich beispielsweise die Suchanfrage nach einem Klangbeispiel des Gesangs des Rotkehlchens leicht beantworten. Ohne die Beschränkung auf den enthaltenen Dateityp Audio wäre die Suche wesentlich zeitaufwendiger.
- Dateigröße: bei *AlltheWeb* läßt sich die Suche auch auf die Dateigröße beschränken (exakt, minimal, maximal). Diese Funktion ist vor allem bei der Suche nach längeren Abhandlungen interessant.

Neben diesen mit *Google* nicht oder nur schwer zu realisierenden Suchanfragen gibt es ein der *Google*-Datenbank eigenes Problem: *Google* erfaßt die gefundenen Dokumente nur bis zu einer Dateigröße von 100 KB. Gerade bei längeren Abhandlungen wird diese Größe jedoch leicht überschritten. Die Suchmaschine *AlltheWeb* bietet eine vollständige Indexierung auch längerer Dokumente.

Vorschläge für die Verbesserung der Suchanfrage

Die „allgemeine“ erweiterte Suche wird wohl in Zukunft nicht mehr stark ausgebaut werden. Information Professionals mögen auf die volle Unterstützung Boole'scher Operatoren und die Möglichkeit der Trunkierung in allen wichtigen Suchmaschinen hoffen; Suchmaschinenbetreiber orientieren sich jedoch vornehmlich an den „normalen“ Nutzern und entwickeln daher eher ihre einfachen Suchinterfaces weiter bzw. unterstützende Funktionen, die diesen intuitiv leiten. Dabei ist eine verblüffende Ähnlichkeit der Suchinterfaces unbestreitbar – letztendlich haben sich viele Suchdienste am Design des *Google*-Interfaces orientiert.

Eine Möglichkeit, dem Benutzer einen intuitiven Zugriff auf die Einschränkung seiner Suchanfrage zu bieten, ist die Clusteranalyse.¹⁴ Angeboten wird dieses Verfahren jedoch bisher von keiner der großen Suchmaschinen¹⁵. Einige kleinere Anbieter wie *Vivisimo* haben sich diese gerade für die Beantwortung unpräziser Suchanfragen wegweisende Technik jedoch zu eigen gemacht.

Weitere Möglichkeiten, den Benutzer zur Präzisierung seiner Suchanfrage zu leiten, sind Vorschläge zur Rechtschreibkorrektur der Suchwörter und der Vorschlag verwandter Suchbegriffe.

Während sich die konventionelle Rechtschreibkorrektur an hinterlegten Wörterbüchern – und damit an einzelnen Wörtern – orientiert, bieten neuere Ansätze

¹⁴ Zum Einsatz der Clusteranalyse bei WWW-Suchmaschinen siehe auch Lewandowski 2001

¹⁵ Eine Ausnahme war *NorthernLight*. Diese Suchmaschine hatte einige Zeit einen der größten Indizes und verband sowohl einfache Suche und Clusteranalyse unter einer intuitiven Oberfläche als auch kostenlose WWW-Inhalte mit Dokumenten der kostenpflichtigen sog. Special Collection (Aufsätze aus Zeitschriften, Agenturmeldungen, Investmentreports). *NorthernLight* wurde im Jahr 2002 von Divine aufgekauft und Ende 2002 / Anfang 2003 nach und nach eingestellt. Die Zukunft dieser einst wegweisenden Suchmaschine ist unklar.

eine Korrektur auf der Basis von Suchanfragen aus der Vergangenheit. Damit lassen sich Begriffe in einen Kontext stellen und beispielsweise auch Fehlschreibweisen in Namen korrigieren. Gibt man zum Beispiel die Suchanfrage „britny spers“ bei *Google* ein, wird aufgrund früherer Fehlschreibungen und der ähnlichen korrekten Schreibweise dem Benutzer diese vorgeschlagen.

Vorschläge verwandter Suchbegriffe bietet unter anderem *AltaVista Prisma*. Aus dem Kontext der in den gefundenen Dokumenten ebenso vorkommenden Begriffe werden weitere Suchwörter angezeigt, die sich leicht entweder in die Suche integrieren oder aber diese ersetzen können.¹⁶

Erweiterte Funktionen

Alle großen Suchmaschinen bieten ihren Nutzern inzwischen mehr als nur die eigentliche Suche nach HTML-Dokumenten und sonstigen Standardformaten. Alle wesentlichen Anbieter bieten inzwischen eine Bilder- und eine Nachrichtensuche an. Neben diesen zusätzlichen Dokumentkollektionen bieten einige Suchmaschinen weitere Features an, die demonstrieren, welche Erweiterungen in Zukunft auch bei den anderen Suchmaschinen anstehen könnten.

Vorreiter auch in diesem Bereich ist *Google*. Dies liegt zum einen an dem Schwerpunkt der kontinuierlichen Entwicklungsarbeit innerhalb des Unternehmens, andererseits an der Einstellung, Entwicklungen schon in einem relativ frühen Stadium den Nutzern über den Bereich *Google Labs* zugänglich zu machen.

Erwähnenswerte Features sind:

- *Google Sets* (<http://labs.Google.com/sets>): nach Eingabe von zwei oder mehr Begriffen werden weitere Begriffe gesucht, die im Kontext der eingegebenen Wörter vorkommen. So lassen sich zum Beispiel durch die Eingabe der Begriffe „*Google*“, „*Alta Vista*“ und „*Yahoo*“ weitere Suchmaschinenanbieter finden. Gerade für die Identifizierung von Firmen, die auf gleichem Gebiet tätig sind oder für die Suche nach weiteren Begriffen für eine Suchanfrage ist dieses Tool geeignet.
- *Google WebQuotes* (<http://labs.Google.com/cgi-bin/webquotes>): nach dem Starten einer normalen Suchanfrage werden in der Trefferliste Kommentare von anderen Websites über die Seiten auf der Trefferliste eingeblendet. Dieses Werkzeug kann nützlich sein, wenn man sich kurz informieren will, wie eine Website oder ein Produkt von anderen eingeschätzt wird. Beispiel wäre die Suchanfrage nach „*Internet Explorer*“, die kurze und durchaus informative Statements zu der entsprechenden Software liefert.
- *Google Answers* (<http://answers.Google.com/answers/main>): Dieser kostenpflichtige Dienst bietet die Möglichkeit, Fragen an von *Google* geprüfte Rechercheure zu stellen. Dabei kann der Preis für die Beantwortung der Suchanfrage vom Suchenden mittels eines Gebots festgelegt werden. Die Qualität und Sinnhaftigkeit dieses Dienstes wurde in Fachkreisen ausführlich diskutiert; einen Erfahrungsbericht bietet West (2002).
- *Teomas* „Link collections from experts and enthusiasts“ (www.teoma.com): bei einer Suche in *Teoma* werden nicht nur die Trefferlisten ausgegeben, sondern auch Verweise auf Seiten, die besonders viele Links auf thematisch verwandte und insgesamt stark verlinkte Seiten enthalten. Diese Funktion ist insbesondere nützlich, um die wesentlichen Quellen zu einem Themengebiet zu identifizieren.

¹⁶ zur Bedeutung und Akzeptanz der Vorschläge ergänzender Suchbegriffe vgl. Belkin 2000

Zukünftige Entwicklungen

Als wichtigste Herausforderungen für die Suchmaschinenbetreiber lassen sich die intuitive Benutzerführung zur Verbesserung unpräziser Suchanfragen, die Einbindung von *Invisible-Web-Content* und insbesondere auch das Finden einer Lösung für den Umgang mit *paid content* ansehen.

Inhalteanbieter sind zunehmend nicht mehr bereit, ihre Inhalte kostenlos an die Nutzer abzugeben. Die meisten Webangebote lassen sich nicht durch Werbung finanzieren. Die Anbieter haben in der Vergangenheit teils auf dieses Refinanzierungsmodell gebaut oder aber keine klare Strategie der Finanzierung entwickelt. Der Trend zu kostenpflichtigen Angeboten wird sich in der nächsten Zeit weiter verstärken, zumal auch auf Nutzerseite zunehmend die Bereitschaft besteht, für sog. *Premium-Content* zu bezahlen (vgl. Verband Deutscher Zeitschriftenverleger 2003, Smart Research 2003).

Wenn nun aber in absehbarer Zeit der Anteil der kostenpflichtigen Inhalte einen bedeutenden Anteil der nachgefragten Inhalte des WWW ausmachen wird, werden die Suchmaschinen ihre bisherige Grenze auf die ein oder andere Weise durchbrechen müssen. Denkbar sind Modelle hybrider Suchmaschinen, die sowohl frei zugängliche Dokumente als auch proprietäre Inhalte indexieren.¹⁷ Denkbar ist auch, daß Inhalte-Anbieter wie teils heute schon für die bevorzugte Indexierung ihrer Dokumente bezahlen oder aber, daß mittels Techniken, die heute als *cloaking* bezeichnet werden, Suchmaschinen einen Volltext zur Indexierung vorgelegt bekommen, während der Nutzer selbst später erst einmal die Aufforderung zur Zahlung angezeigt bekommt¹⁸, bevor er den Volltext sehen kann. Der Vorteil für den Nutzer läge darin, daß er mittels der Suchmaschinen weiter die Volltexte durchsuchen könnte, während er sonst nur eine Zusammenfassung (bzw. den Teaser) eines Texts mittels der Suchmaschinen finden könnte.

Für die Suchmaschinenbetreiber bedeutet es schlicht eine Frage des Überlebens: wenn sie nicht mehr die von den Nutzern gewünschten Inhalte indexieren können, so werden sich diese von den für sie untauglichen Suchwerkzeugen abwenden.

¹⁷ Eine solche Hybridlösung bot etwa bis zu ihrer Einstellung die Suchmaschine *NorthernLight* an. Eine Hybridlösung für wissenschaftliche Webinhalte und wissenschaftliche Zeitschriftenaufsätze bietet www.scirus.com.

¹⁸ *Cloaking* ist „the practice of delivering content based on the IP address of the client“ (Search Engine Directory 2003)

Literatur

Bates 2003

Bates, Mary Ellen: Online Spotlight: HotBot. Online 27(2003)2, 80

Belkin 2000

Belkin, Nicholas J.: Helping People Find What They Don't Know. Communications of the ACM 43(2000)8, 58-61

Bergman 2001

Bergman, Michael K.: The Deep Web: Surfacing Hidden Value (Bright Planet Deep Content White Paper; Spetember 2001).
<http://www.brightplanet.com/deepcontent/tutorials/DeepWeb/deepwebwhitepaper.pdf> [18.3.2003]

Gelernter 2003

Gelernter, Judith: At the limits of Google: Specialized Search Engines. Searcher 11(2003)1, 26-31

Griesbaum et al. 2002

Joachim Griesbaum, Marc Rittberger, Bernard Bekavac, Deutsche Suchmaschinen im Vergleich: AltaVista.de, Fireball.de, Google.de und Lycos.de, in: Rainer Hammwöhner, Christian Wolff, Christa Womser-Hacker (Hg), Information und Mobilität, Optimierung und Vermeidung von Mobilität durch Information, Proceedings des 8. Internationalen Symposiums für Informationswissenschaft, Konstanz 2002, S.201-223

Kuhlen 1999

Kuhlen, Rainer: Die Konsequenzen von Informationsassistenten: was bedeutet informationelle Autonomie oder wie kann Vertrauen in elektronische Dienste in offenen Informationsmärkten gesichert werden? Frankfurt am Main 1999

Lewandowski 2001

Lewandowski, Dirk: „Find what I mean not what I say“: Neuere Ansätze zur Qualifizierung von Suchmaschinen-Ergebnissen. BuB 53(2001)6/7, 381-386

Lewandowski 2001

Lewandowski, Dirk: Alles nur noch Google? – Entwicklungen im Bereich der WWW-Suchmaschinen. BuB 54(2002)9, 558-561

Notess 2002

Notes, Greg R.: Search Engine Statistics: Freshness Showdown.
<http://www.searchengineshowdown.com/stats/freshness.shtml> [18.3.2003]

Notess 2003

Notess, Greg R.: Search Engine Statistics: Database Total Size Estimates.
<http://www.searchengineshowdown.com/stats/sizeest.shtml> [18.3.2003]

Search Engine Directory 2003

Le Roux, André [Hrsg.]: Search Engine Directory.
<http://www.searchenginedictionary.com> [18.3.2003]

Smart Research 2003

Paid Content im eCommerce - Eine Studie zur Zahlungsbereitschaft für kostenpflichtige Internetinhalte. Köln 2003

Stiftung Warentest 2003

Noch liegt *Google* vorn – Test Internet-Suchmaschinen. Test (2003)2, 38-41

Sullivan 2003

Sullivan, Danny: Overture To Buy FAST Web Search Division. Search Day Number 472 (26.2.2003).
<http://www.searchenginewatch.com/searchday/03/sd0226-overfast.html>
[18.3.2003]

Verband Deutscher Zeitschriftenverleger 2003

Verband Deutscher Zeitschriftenverleger [Hrsg.]: „Paid Content – Der Markt für Online Inhalte“ Berlin 2003

West 2002

West, Jessamyn: Information for Sale: My Experience With *Google* Answers. Searcher 10(2002)9, 14-19