

Paradies für Spürnasen

Sowohl für den Surfer als auch den Seitenbetreiber ist es hilfreich, die Arbeitsweise von Suchmaschinen, Verzeichnissen und Portalen zu kennen



Fast alle Surfer – rund 85 Prozent – verwenden Suchmaschinen und Portale, um an Informationen zu gelangen. Und 90 Prozent dieser Suchmaschinen-Benutzer erreicht man als Webmaster mit fünf Suchmaschinen. Daher lautet die Devise nicht, wie viele Pseudospezialisten propagieren, in möglichst vielen Suchmaschinen präsent zu sein. Vielmehr ist es wichtig, bei den Top-Playern eine gute Position zu erhaschen – und zu halten.

Suchmaschinen: Der Content zählt

Suchmaschinen sind erwachsen geworden. Zu Beginn des Internet-Booms gab es noch eine Menge Tricks, mit denen sich Webseiten, die thematisch nichts mit den Suchbegriffen zu tun haben, auf die begehrten Plätze bei den Suchergebnissen wichtiger Engines bringen ließen. Das ist mittlerweile schwierig geworden und wird zum Teil mit dem Ausschluss der Seite aus dem Suchverzeichnis bestraft. Besonders beliebt war das Keyword-Spamming, also die Anhäufung von Suchbegriffen als weißer Text auf weißem Grund – dafür gibt es heute bei Google

und Co. die rote Karte. Alle Benutzer haben das Recht, die entsprechenden Suchmaschinenbetreiber auf etwaige manipulierte Einträge hinzuweisen. Und da die Konkurrenz im Online-Business hart geworden ist, liegen ständig Neider auf der Lauer, die nur darauf warten, eine Website anzuschwärzen. Ein Grund, warum Suchmaschinen daran interessiert sind, nur qualifizierte Ergebnisse zu liefern, ist ihre Reputation. So hat sich die Suchmaschine Google binnen weniger Monate nach dem Start als unangefochtener Spitzenreiter hervorgetan, unter anderem, indem sie als erste Engine besonderen Wert auf die Qualität und nicht nur auf die Quantität der Suchergebnisse gelegt hat. Google filterte von Anfang an Seiten aus, bei denen der Autor versuchte, mithilfe billiger Tricks an die Spitze zu gelangen. Derzeit werden die einzelnen Seiten anhand von zwei Hauptkriterien bewertet: Position sowie Dichte der Stichwörter und Stichwortkombinationen. Das gilt für alle Suchmaschinen, die sich nur darin unterscheiden, dass die einen mehr Wert auf den dargestellten Text der Seite legen und die anderen dem Titel der Seite oder verschiedenen Auszeichnungen wie Links oder Überschriften mehr Bedeutung beimessen.

Spider: So werden Sie gefunden

Bei der Anmeldung einer Website ist es normalerweise nicht nötig, jedes einzelne Dokument registrieren zu lassen. Nachdem der Webmaster die Seite im entsprechenden Anmeldeformular eingetragen hat, wird sie in einer Datenbank auf einen Stapel gelegt. Von diesem Stapel nimmt sich der Suchmaschinenroboter – Spider oder Crawler genannt – jeweils den untersten URL, lädt das entsprechende Dokument und folgt den Links innerhalb der Website. Dieser Vorgang dauert von wenigen Tagen bis hin zu einigen Monaten. Dabei ist es in der Regel kontraproduktiv, die Seite öfter anzumelden, da sie schlimmstenfalls bei erneuter Anmeldung aus dem unteren Teil des Stapels entfernt wird.

Sind die einzelnen Seiten dann im Verzeichnis der Suchmaschine aufgenommen, entscheidet ein weiterer Algorithmus darüber, wie wichtig die Seite ist und in welcher Frequenz sie wieder vom Crawler besucht werden soll. Keine Suchmaschine gibt den genauen Algorithmus preis. Allgemein bekannt ist jedoch, dass einerseits Frequenz und Position der wichtigen Schlagwörter und andererseits die so genannte Link Popularity eine wichtige Rolle bei der Bewertung spielen. Bei Fre-



quenz und Position geht es darum, wie häufig und wo innerhalb der Webseiten die Schlüsselwörter oder Schlüsselwortkombinationen zu finden sind. Auf die Link Popularity, die insbesondere Google für seine Bewertung heranzieht, hat der Webmaster nur indirekten Zugriff. Sie steht für Anzahl und Qualität von Verlinkungen von anderen bekannten Seiten auf die entsprechende Webseite, wobei ein Link einer wichtigen Seite, wie beispielsweise Yahoo, mehr ins Gewicht fällt als tausend Einträge in unbekanntenen Link-Listen.

**Verzeichnisse:
Von Mensch zu Mensch**

Noch sind Crawler und Roboter nicht so weit, dass sie Seiten linguistisch perfekt analysieren können. Daher lässt sich eine Seite mit Tricks bei den Suchergebnissen nach oben pushen. Anders ist dies bei Verzeichnissen. Das wohl bekannteste und in Deutschland wichtigste ist Yahoo. Eingeteilt in Ober- und Unterkategorien finden sich hier thematisch ähnliche Seiten. In einem Verzeichnis kann man nach Stichwörtern suchen oder über die Hierarchie zum gewünschten Themenbereich gelangen. Vor allem weniger versierte Benutzer gehen gerne den letztgenannten Weg, weil es ihnen oft schwer fällt, Suchwortkombinationen zu formulieren. Wer „Auto“ eingibt, um bei Google nach „gebrauchten Mini-Cabrios in

Bayern“ zu suchen, wird in der Suchmaschine kaum fündig. In einem Verzeichnis wie Yahoo hingegen wird der Laie Schritt für Schritt durch die Kategorien und Unterbereiche geführt.

Die einzelnen Websites werden von Redakteuren besucht. Ob eine Seite aufgenommen wird, entscheidet der Redakteur anhand seiner Vorgaben, ebenso über die Kategorie, in die der Webauftritt fällt.

Welche Seiten von Redakteuren besucht werden, hängt von unterschiedlichen Kriterien ab. So durchsuchen sie ständig das Internet nach aktuellen Mainstream-Themen. Diese Seiten werden dann automatisch dem Index hinzugefügt. Zum anderen kann der Webmaster die Seite anmelden. Allerdings ist die Anmeldung oft nur für private Seiten kostenlos. Für kommerzielle Seiten verlangen etwa Yahoo oder Looksmart einmalig eine Aufnahmegebühr und andere Verzeichnisse wie Allesklar, Web.de einen jährlichen Obolus.

In diesem Zusammenhang wichtig ist das Open Directory Project (www.dmoz.de). Dieses von der Firma Netscape initiierte Projekt dient vielen Suchmaschinen, wie beispielsweise Google, AOL und Dino-Online, als Lieferant für den Verzeichnisbereich. DMOZ selbst ist ein Service, der nicht auf Profit ausgerichtet ist. Jeder kann hier Redakteur einer Rubrik werden – allerdings ohne Entgelt. Dies hat Vor-

Zukunftsmusik

→ Der Trend bei der Analyse von Inhalten geht in Richtung Verständnis des Inhalts. So entwickelt sich die linguistische Forschung immer weiter. Auf diese Weise beginnen Algorithmen langsam, die Inhalte der einzelnen Webseiten zu analysieren und die Site im Kontext zu sehen. Dennoch ist nicht damit zu rechnen, dass die Dichte der Schlüsselwörter – derzeit das A und O bei der Optimierung – in Zukunft gänzlich an Relevanz verliert. Vielmehr werden sich Seiten hervortun, bei denen es im gesamten Auftritt um ein und dasselbe Thema geht und die es schaffen, dies auch verständlich zu vermitteln.

und Nachteile. Von Vorteil ist, dass der Eintrag kostenlos ist. Die Kehrseite ist, dass es kein Kontrollorgan gibt und somit der Subjektivität Tür und Tor geöffnet ist.

Portale: Unter seinesgleichen

Obwohl – abgesehen vom Branchenbuch Bellnet – Themenportale nicht zu den Top-10-Suchmaschinen gehören, bringen sie doch sehr qualifizierten Traffic auf die Webseite. Der Grund dafür liegt auf der Hand. Wenn sich ein Surfer über einen Internet-Shop-Katalog beispielsweise über das Thema Tee informiert, kann man ein grundsätzliches Kaufinteresse voraussetzen. Insofern ist natürlich ein Besucher eines Shopping-Portals eher ein potenzieller Käufer als jemand, der bei Google nach dem Begriff „Tee“ suchen lässt. Zudem steigert ein



Eintrag in einem solchen Portal die Wichtigkeit des URL in den Augen der Suchmaschinen, was zu einer höheren Link Popularity (bei Google Pagerank genannt) und damit zu einem höheren Ranking führt. Welche Portale besonders empfehlenswert sind, kann man leicht überprüfen, indem man bei einer Suchmaschine nach den entsprechenden Begriffen, wie beispielsweise „Shop Verzeichnis“, sucht. Alternativ gibt es bei den Verzeichnissen des Open Directory Project respektive bei Dino-Online und Yahoo je ein ganzes Verzeichnis nur für Shop-Portale. Auch die Themenportale werden meist von Redakteuren gewartet und enthalten nur Seiten, die thematisch passen.



Wichtig: Bei den meisten Themenportalen können Webseiten kostenlos angemeldet werden. Die einzige Gefahr liegt darin, dass der Redakteur die Seite als nicht würdig erachtet und sie somit nicht im Index erscheint. Es gibt aber auch schwarze Schafe, die nur innerhalb der AGBs vermerken, dass mit einer Listing Kosten für vermittelte Besucher oder auch monatliche Gebühren fällig werden. Daher ist es immer ratsam, die AGBs vor der Anmeldung zu lesen oder zumindest nach Begriffen wie „Euro“ oder „€“ zu durchsuchen.

Quellen der Information
Meinte zu Anfang jeder Suchdienst, seine eigene Datenbank aufbauen zu müssen, setzt man heute auf Synergieeffekte. So teilen sich viele Suchmaschinen den Verzeichnisbereich, der wegen der nötigen Manpower teuer ist. Google als Nummer eins sowie AOL und Altavista verwenden die Daten des Open Directory Project. Allerdings wird der Directory-Bereich etwas stiefmütterlich

behandelt. So kann es mehrere Monate dauern, bis ein im DMOZ aufgenommenen Eintrag auch im Verzeichnis der Suchmaschine sichtbar ist. Im Gegensatz dazu leisten sich Allesklar.de, Web.de und Yahoo eigene Redaktionen. Die Aufnahme ist für kommerzielle Seiten kostenpflichtig, dafür aber schnell. Die Aufnahme privater Auftritte dauert – wenn sie überhaupt gelistet werden – oft Monate. Besonders krass zeigt sich die Synergie bei den tatsächlichen Suchergebnissen der Crawler-basierenden Datenbanken. Gab es noch bis vor einem Jahr knapp ein halbes Dutzend Informationsquellen, hat sich das Feld nun auf zwei Datenbanken reduziert: Google und Inktomi. So finden sich die Ergebnisse von Google auch bei AOL und T-Online wieder und MSN, Altavista und Web.de beziehen ihre Informationen von der Inktomi-Datenbank. Diese Datenbank wurde im Dezember 2003 von Yahoo gekauft. Für den Administrator einer Website hat das den Vorteil, dass er mit der

Oft getarnt: Bezahlte Suchergebnisse



► **Nix klar bei Allesklar**
Beim Verzeichnis Allesklar.de sind die bezahlten Einträge (hier grün und gelb markiert) kaum von den echten Treffern (hier blau) zu unterscheiden



▲ **Lernfähig** Auch bei T-Online geht die Kennzeichnung der kommerziellen Treffer (hier grün markiert) im Wust der Informationen unter

Eintragung in zwei Datenbanken bei nahezu 99 Prozent aller Suchanfragen berücksichtigt wird. Doch obwohl beispielsweise MSN und Altavista dieselbe Datenbasis verwenden, erhalten Besucher bei derselben Suchanfrage unterschiedliche Ergebnisse. Das liegt daran, dass trotz einheitlicher Daten unterschiedliche Verfahren (Algorithmen) verwendet werden, um Suchanfragen zu bearbeiten. Anders ist dies bei Google und den darauf zugreifenden Diensten wie T-Online oder AOL. Diese übernehmen nicht nur die vom Googlebot gesammelten Informationen, sondern auch die Ergebnisse. Einen sehr guten Überblick über die Verstrickungen der einzelnen Anbieter findet man auf der interaktiven Seite www.bruceclay.com. In einem Diagramm wird dort die Herkunft der Informationen interaktiv und transparent dargestellt.

Banner im Treffer-Look

Selten sind gekaufte Ergebnisse so gut sichtbar wie bei Google. Hier ist die Werbung am rechten und oberen Rand farblich hervorgehoben und ganz klar und eindeutig von den eigentlichen Suchergebnissen getrennt. Viele Suchmaschinen stellen bezahlte Einträge allerdings innerhalb der Ergebnisse ihrer Suchdatenbank dar. Besonders irreführend ist Allesklar.de. Während die Empfehlungen am rechten Rand noch deutlich als Werbung erkennbar sind, scheinen die Einträge im eigentlichen Ergebnissenfenster sämtlich Treffer der Suche zu sein. Weit gefehlt! Wer das graue Kleingedruckte liest, erfährt, dass im ersten Bereich „gesponserte Webseiten“

Marktanteile Suchmaschinen

Suchmaschine	URL	Marktanteil
Google	www.google.de	76,5 %
Yahoo	www.yahoo.de	5,7 %
MSN Web-Suche	www.msn.de	4,8 %
AOL Suche	www.aol.de	2,3 %
T-Online	www.t-online.de	2,3 %
WEB.DE	www.web.de	1,7 %
Altavista	www.altavista.de	1,0 %
Lycos	www.lycos.de	0,9 %
Fireball	www.fireball.de	0,4 %
Sonstige		4,4 %

QUELLE: www.webhits.de

– also Werbung – angezeigt wird. Und auch danach erscheinen keine Suchtreffer, sondern Einträge aus dem Directory, die für den Anbieter ebenfalls kostenpflichtig sind.

Auch bei T-Online ist nicht sofort ersichtlich, dass die ersten Einträge keine Fundstellen sind, sondern Werbung. Das sieht man nur an dem unscheinbaren Sponsored Link, der in der textlastigen Darstellung nahezu untergeht.

Dass es nicht so penetrant sein muss, beweist beispielsweise MSN. Früher genauso versteckt wie bei Allesklar.de und T-Online, hebt sich die Werbung bei MSN mittlerweile ebenso gut wie bei Google von den „echten“ Ergebnissen ab.

Ingo Böhme/pat

In Zusammenarbeit mit REDTEC IT-Fachinformationen

Die Nummer eins im Suchgeschäft



Info Google ist beliebt

➔ Google hat sich innerhalb weniger Monate zur weltweit beliebtesten Suchmaschine entwickelt. Bei Surfern ist sie begehrt, weil sie einerseits die qualitativ besten Ergebnisse liefert. Zum anderen ist sie auch mit die schnellste Engine, die es heutzutage im Internet gibt. Das bezieht sich sowohl auf die samt aller Bilder knapp 14 KByte große und simpel aufgebaute Suchseite als auch auf die Zeit, die verstreicht, bis alle Ergebnisse einer Suche angezeigt sind.

Von Webmastern wird Google geschätzt, weil sie nahezu die einzige Suchmaschine ist, die kein Geheimnis daraus macht, wie Seiten indiziert und bewertet werden (www.google.com/technology). Zudem wird eine neue Seite binnen weniger Tage vom Googlebot besucht und kommt somit auch schnell in den Ergebnissen vor. Und wenn der Inhalt der Seiten stimmt, kehrt der Bot in kurzen Intervallen wieder auf die Seite zurück. Das wiederum hat zur Folge, dass Optimierungsversuche sich schon nach kurzer Zeit in der Praxis widerspiegeln.

◀ **Überzeugend** Google trennt bezahlte Inhalte (oben und rechts) vorbildlich von solchen der DMOZ-Redaktion (gelb) und Crawler-Suchergebnissen