



Jo Bager

Turbolift ins Ungewisse

Manipulationsmaschen bei Suchmaschinen

Seit es Suchmaschinen gibt, versuchen findige Webmaster ihre Sites mit allerlei Tricks in die Spitzenpositionen der Trefferlisten zu mogeln. Suchmaschinen quittieren solche Versuche durch ein schlechtes Ranking oder die Verbannung aus dem Index – sofern sie sie bemerken. Mitunter bestrafen sie auch unbescholtene Webmaster, die unwissentlich und ohne böse Absicht gegen die Spielregeln verstoßen.

Als Mitte der 90er Jahre die ersten Suchmaschinen an den Start gingen, war schnell klar: Schafft ein Webmaster es, dass Suchmaschinen die eigene Site an prominenter Stelle der Ergebnislisten präsentieren, so ist seiner Online-Präsenz ein breiter Besucherzustrom sicher, der sich in bare Münze verwandeln lässt.

So entbrannte ein bis heute andauernder Wettstreit zwischen Suchmaschinenbetreibern und Suchmaschinenoptimierern um die Kontrolle der Ergebnislisten [1]. Die Manipulationsmöglichkeiten der Suchmaschinen-trickser und die Gegenmaßnah-

men der Suchdienste zu kennen, heißt für Webmaster, die mit sauberen Mitteln arbeiten wollen, mögliche Fehlerquellen vermeiden zu können.

In den Anfängen ließen sich Suchdienste wie AltaVista oder HotBot durch so genanntes Keyword Stuffing dazu bringen, eine Seite hochzupushen: Der gewünschte Suchbegriff wurde einfach immer und immer wieder in eine Seite hineingeschrieben: In den Titel, in die Meta-Tags, in den Body, überallhin, Hauptsache oft. Die Seite mit den meisten Schlüsselbegriffen gewann.

Heutzutage werten Suchdienste Seiten mit einer zu hohen

Keyword-Dichte ab, also einem zu hohen Anteil eines Schlüsselbegriffs im Vergleich zum Rest des Textes. Insbesondere versteckte Texte und Links sehen Suchmaschinen als Foul an. Dazu gehören zum Beispiel Elemente, die in der Hintergrundfarbe der HTML-Seite verfasst sind, um den Augen der Surfer verborgen zu bleiben, von den Suchmaschinen-Robots aber registriert zu werden. Wegen der vielen Möglichkeiten, über geschachtelte Tabellenkonstruktionen, Layer und Cascading Stylesheet eine Seite zu gestalten, sind solche Manipulationen automatisch jedoch schwer zu erkennen.

Vollgestopft

Das Vollstopfen mit Schlüsselwörtern wirkt auch heute noch negativ nach. So messen die Suchdienste dem ursprünglich sinnvollen Meta-Keyword-Tag, mit dem Webmaster die relevanten Schlüsselbegriffe ihrer

Sites beschreiben, wenig Wert bei.

Der seinerzeitige Suchmaschinen-Primus AltaVista wurde durch Keyword-bepackten Spam derart vollgestopft, dass er Ende 1997 kaum mehr zu gebrauchen war – ein Problem, von dem sich AltaVista nie mehr so richtig erholen konnte, denn 1998 trat ein anderes Unternehmen auf den Plan, das schnell zur Nummer Eins avancieren sollte: Google.

Eine der Neuerungen von Google war ein Ranking-Algorithmus namens Page Rank, der lange Zeit als immun gegen Manipulationen galt [2]. Der Page Rank einer Site berechnet sich aus der Anzahl der Links, die auf die Site verweisen und den Page Ranks (und damit der Popularität) der verweisenden Sites. Sites, die von vielen, wichtigen Sites verlinkt werden, sind ebenfalls wichtig und tauchen in den Spitzenpositionen der Trefferlisten auf. Allgemein spricht man von Link-Popularität statt nur vom Page Rank, denn alle größeren Such-

maschinen untersuchen mittlerweile die Verweisstruktur bei der Berechnung von Rankings. Neben der Zahl der Links spielt auch der Ankertext der Links eine Rolle. Das verlinkte Wort wird dem Zieldokument des Links zugeschlagen und erhöht dessen Relevanz in Bezug auf dieses Wort.

Im Jahre 2001 entdeckte Adam Mathes, dass sich diese Eigenart des Google-Ranking-Algorithmus für Manipulationen ausnutzen ließ. Er hatte die Worte „talentless hack“ benutzt, um damit einen Verweis auf die Webseite eines Freundes zu kennzeichnen. Nachdem er andere Webmaster aufgefordert hatte, mit demselben Text auf dieselbe Seite zu verweisen, erscheint wenig später die Seite des Freundes bei Google auf Platz 1, wenn der Benutzer „talentless hack“ als Suchbegriff eingab.

Dieser Trick, der „Google Bombing“ getauft wurde, wird bis heute gelegentlich wiederholt. Insbesondere in der Blogger-Welt ist diese Masche recht populär. So führt nach einer Google-Bombe der Begriff „miserable failure“ nach wie vor zur Biographie von US-Präsident George Bush.

Link-Fabriken

Ein großes Problem stellen Google Bombs aber nicht dar, da es nur in einzelnen Fällen benutzt wird, die auch publik werden. Der Beweis, dass sich auch der Page Rank manipulieren lässt, führte allerdings dazu, dass alle Dämme brachen. So genannte Link-Farmen wurden aus dem Boden gestampft. Auf dutzenden oder hunderten von Domains generieren dabei Pro-

gramme aus umfangreichen Stichwortlisten tausende untereinander verlinkter Seiten.

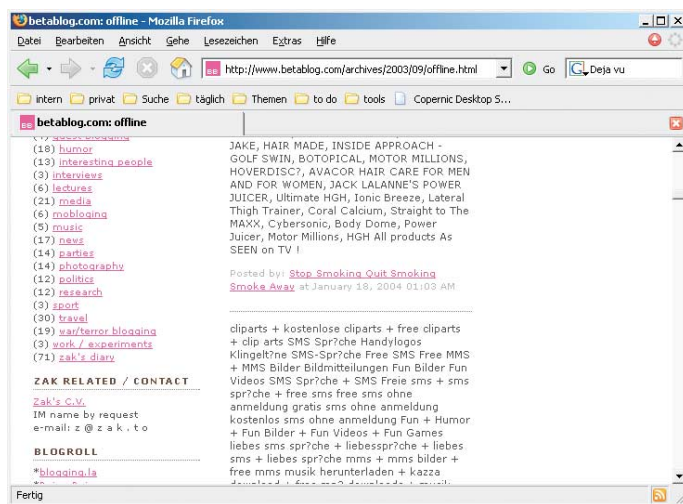
Mitte 2003 haben Link-Farmen zu einer wahren Inflation von Spam-Ergebnissen bei Google geführt. Mittlerweile hat Google das Problem besser im Griff – und versucht Trickereien mit der Verlinkung aufzuspüren und zu bestrafen. So warnt Google Webmaster in seinen Richtlinien davor, an Linktauschprogrammen teilzunehmen, die nur den Zweck verfolgen, das Ranking oder den Page Rank zu verbessern [3].

Links zu „Spammern“ oder „schlechten Nachbarn“ sollte der Webmaster vermeiden, weil mit deren Abstufung im Ranking auch das der verweisenden Sites leiden könnte. Was genau Google unter Spammern oder schlechten Nachbarn versteht, lässt der Suchmaschinenbetreiber aber im Unklaren – wie so vieles in den Richtlinien. Ein Webmaster sollte daher genau überlegen, mit welchen Sites er seine Präsenzen verlinkt – egal, wie reizvoll ein Verweis von einer Site mit hohem Page Rank sein mag.

Kuckuckseier

Seit 2003 ist es unter Suchmaschinen-Spammern Mode, nicht eigene Link-Farmen einzusetzen, sondern fremde Web-Auftritte zu missbrauchen, um Links auf die eigenen Präsenzen zu setzen, zum Beispiel in Foren oder Gästebüchern – überall, wo sich Besucher einer Site verewigen dürfen. Deshalb wird dies auch als Gästebuch- oder Kommentar-Spam bezeichnet.

Um dem Kommentarspam das Wasser abzugraben, hat



Beim Gästebuch-Spam, wie hier in einem Blog, versuchen schwarze Schafe fremde Sites einzuspannen, um die Link-Popularität des eigenen Online-Auftritts zu verbessern.

Google gemeinsam mit MSN und Yahoo eine technische Lösung vorgeschlagen. Links, die jeder Internutzer anlegen kann, sollen künftig von Suchmaschinen beim Berechnen der Link-Popularität nicht mehr beachtet werden. Technisch wird das möglich durch die von der Blog-Software vorgenommene Ergänzung des Link-Tags um das Attribut:

```
rel="nofollow"
```

Viele Betreiber von Blog-Diensten setzen dieses Attribut bereits ein. Daneben hat die Netzgemeinde weitere Mittel gegen Kommentarspammer hervorgebracht, zum Beispiel Captchas. Das sind Codes, die von einem Script als Bild ausgegeben werden, und vom Besucher in ein Feld übertragen werden müssen, bevor dieser einen Kommentar oder einen Forumsbeitrag verfassen kann. Die Skripte der Suchmaschinen-Spammer scheitern daran.

Ein Webmaster, der seine Blog- oder Foren-Software selbst betreibt, sollte dringend ein Mittel gegen Kommentarspam einbauen, um Spammer von seinen Seiten fern zu halten. Sonst könnte seine Site möglicherweise von Google für unerwünschte Links abgestraft werden.

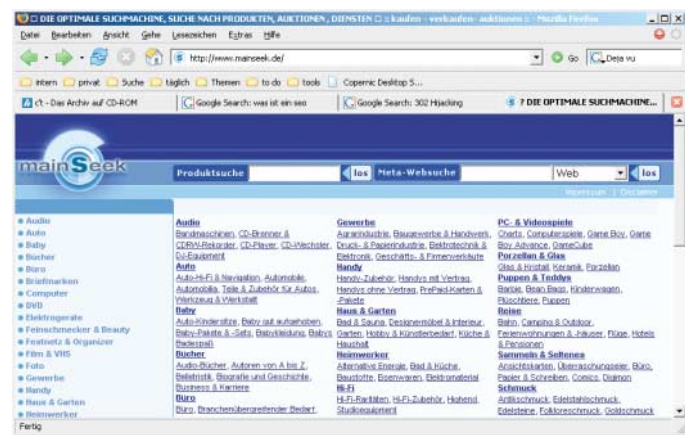
Neben dem Aufbau von Sites mit künstlich hochgezüchtetem Page Rank hat sich auch ein regelrechter Handel mit Domains entwickelt, die einen hohen Page Rank besitzen. Wer zum Beispiel auf eBay mit dem Suchbegriff Pagerank sucht, findet

dutzende Sites mit einem (angeblich) hohen Page Rank. Vor dem Erwerb einer solchen Site sei eindringlich gewarnt. Man weiß nie, mit welchen Mitteln diese Sites ihre hohe Einstufung erhalten haben, und ob sie nicht in der nächsten Woche schon abgestraft werden. Zudem gibt es Mittel, die Anzeige eines Page Rank zu manipulieren. Eine vermeintlich hoch gerankte Seite erweist sich bei genauerem Hinsehen oft als Fake.

Wo ist das Suchwort?

Fast so alt wie das Keyword Stufing sind die verschiedenen Methoden, Suchmaschinen-Robots andere Inhalte zu präsentieren als Besuchern, um die Ranking-Mechanismen auszuhebeln [4]. Brückenseiten (Doorway Pages) zum Beispiel sind für bestimmte Suchbegriffe optimierte Seiten. Ruft ein Browser die Seite auf, wird er, zum Beispiel per JavaScript-Befehl, auf einen anderen Inhalt geleitet.

Suchmaschinen-Robots können nur einfache, nicht aber komplexe JavaScript-Befehle auswerten, erhalten also die optimierte Seite. Beim Cloaking wird unter derselben URL dem Suchmaschinen-Robot ein anderer Inhalt angezeigt als dem Surfer. Wenn eine Suchmaschine eine Seite verlinkt, die nicht das Suchwort enthält, so ist dies ein starker Hinweis auf eine Brückenseite oder Cloaking (oder eine Link-Farm, die sich den „Google Bombing“-Effekt zu Nutze macht).



Fast schon ein Klassiker: Link-Farmen wie Mainseek sollten das Ranking-Verfahren von Google austricksen.

Brückenseiten lassen sich nicht immer vermeiden, insbesondere wenn es darum geht, datenbankgenerierte Seiten für Suchmaschinen sichtbar zu machen. Dafür sollte man aber einen Suchmaschinen-Dienstleister beauftragen, der weiß, wie man solche Seiten baut, ohne dass die Suchdienste es ahnden.

Déjà vu

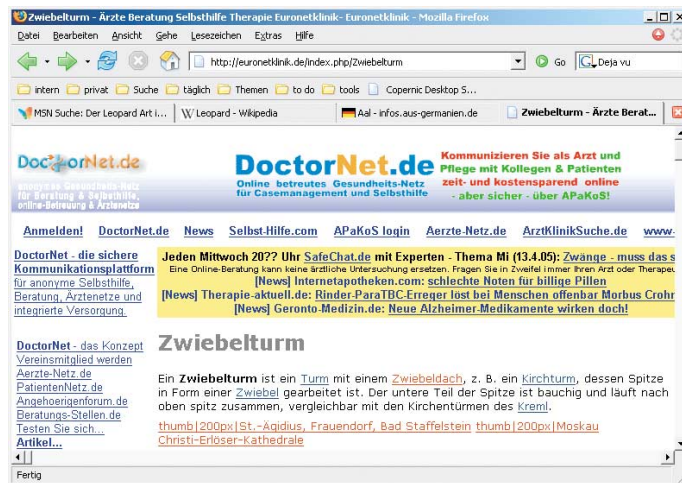
Heute gibt es ausgezeichnete kostenlose Informationen im Netz, die noch dazu jedermann in seine Seiten einbauen darf: Das Web-Verzeichnis des Open Directory Project wird schon auf hunderten anderer Seiten verwendet, diverse Newsticker liefern aktuelle Schlagzeilen, die Wikipedia zigtausende hochwertiger Lexikonbeiträge.

Wer im Internet Geld verdienen will und selber keine Produkte zu verkaufen hat, betreibt Werbeflächen für die großen Werbenetzwerke Google Adwords, Overture und Espotting oder bindet Provisionslinks von allen möglichen Shops ein. Freiberufliche Suchmaschinenoptimierer leben davon, Content und Werbung so elegant zu vermischen, dass der normale Nutzer im Idealfall gar nichts davon bemerkt.

Hinter den Trefferzahlen von Google und Co. verbirgt sich auch eine riesige Zahl von hochprofessionell gemachten Seiten mit wiedergekäuten Inhalten. Und genau solche Dubletten, Fachjargon: duplicate content, hat Google derzeit im Visier [3]: „Don't create multiple pages, subdomains, or domains with substantially duplicate content.“

Die Bestrafung von doppelem Content zeigt, wie schnell sich die Spielregeln im Suchmaschinenoptimierer-Gewerbe ändern können. Noch vor kurzer Zeit galt es als Tipp, mehrere Domains auf denselben Inhalt abzubilden, also zum Beispiel unter www.zuchtbetrieb-danuber.de und www.hommingberger-gepardenforelle.de identische Websites zu betreiben.

Einige Provider bilden bei Hosting-Paketen, die mehrere Domains umfassen, per Default sämtliche Domains auf dieselben Dateien auf dem Webserver ab. Damit dürften derzeit tausende Domains von der Löschung aus dem Google-Index oder zumindest von der Abstufung bedroht sein. Es wurde in



Was hat ein Zwiebelturm mit dem Thema Gesundheit zu tun? Nichts, aber Wikipedia-Artikel wie dieser verhelfen der Site zu einem besseren Ranking.

einschlägigen Foren auch bereits über Fälle berichtet, in denen derselbe Inhalt unter www.domainname.de und domainname.de (ohne www.) zur Bestrafung führte.

Nichtsdestotrotz kann man sich auch heute noch frei verfügbarer Quellen bedienen. Wenn man aber zum Beispiel einen Wikipedia-Artikel nutzt, so sollte man ihn auf keinen Fall 1:1 kopieren. Statt dessen empfiehlt es sich, wie bei einem Schulaufsatz, den Artikel durchzulesen und mit eigenen Worten wiederzugeben. Noch besser, wenn auch mit mehr Arbeit verbunden, ist es jedoch, eigenen Inhalt zu schaffen.

Rausgekickt

Das brutalstmögliche Mittel im Kampf um die vorderen Ränge bei Google und Co. besteht darin, Konkurrenz-Sites aus dem Index zu tilgen. Beim Suchmaschinen-Optimierer-Wettbewerb Nigritude Ultramarine wurde

dieses des Öfteren praktiziert [5]. So scheint Google bei Dubletten das Original anhand des Page Ranks zu bestimmen: Die Site mit dem höchsten Page Rank ist das Original, alle anderen Websites sind Kopien und werden heruntergestuft. Wenn allerdings ein Suchmaschinenoptimierer kopierte Inhalte auf einer Link-Farm mit hohem Page Rank unterbringt, so kann es passieren, dass diese fälschlicherweise als Original angesehen und der wahre Urheber abgestraft wird.

Eine weitere Technik, das so genannte 302 Hijacking, ermöglicht es Angreifern, dass seine Seiten statt denen des Opfers in den Suchergebnissen bei Google erscheinen. Wer einen Verstoß gegen die Richtlinien der Suchmaschinen bemerkt, sollte ihn melden. Alle Suchmaschinen-Betreiber bieten Formulare an, mit denen Webmaster Spam melden können. Das Google-Formular findet sich zum Beispiel unter [6]. Man sollte sich bei einer solchen

Meldung allerdings keine Hoffnung auf eine schnelle Behebung spezifischer Probleme machen. Google nimmt solche Hinweise zum Anlass, die Probleme algorithmisch zu lösen, was einige Zeit dauern kann.

Fazit

Es gibt sie, die Expresslifte in die Top-Positionen bei Google, Yahoo, MSN und Co., – zum Beispiel in Form von Domains mit hoher Link-Popularität für 200 Dollar bei eBay. Allerdings gibt es keine Gewähr, dass die Website für länger als ein paar Tage auf der Sonnenseite bleibt, im Gegenteil. Die Suchdienstleister rüsten ihre Manipulations-Radare immer weiter auf, sodass die Web-Präsenz nach kurzer Zeit wieder auf einem schlechten Platz in den Trefferlisten liegen dürfte und der Webmaster wieder am Anfang steht. Dauerhafter Erfolg ist nur auf dem Weg der kontinuierlichen Optimierung gegeben. (jo)

Literatur

- [1] SEO Methodologies, A brief history of our industry, Abriss von Brett Tabke, www.webmasterworld.com/forum5/1008.htm
- [2] Originalpaper zu Page Rank, <http://www.db.stanford.edu/~backrub/google.html>
- [3] Googles Webmaster Guidelines, www.google.com/webmasters/guidelines.html
- [4] www.linksandlaw.de/suchmaschinen-leitfaden-zur-Manipulation-von-Suchergebnissen.htm
- [5] FAQ zum Suchmaschinenoptimierungs-Wettbewerb Nigritude Ultramarine, www.nigritudeultramaries.com
- [6] Googles Formular für Spam-Meldungen, www.google.com/contact/spamreport.html

Jagd frei auf die Hommingberger Gepardenforelle!

c't ruft auf zum Wettstreit der Suchmaschinenoptimierer. Ziel ist es, für den Begriff „Hommingberger Gepardenforelle“ eine Top-Position in den Google.de-, Yahoo.de-, MSN.de- und Seekport.de-Ergebnislisten zu ergattern. Wir beobachten während des Wettbewerbs die Positionierungen der besten Sites in den Trefferlisten. An zwei Stichtagen, am 15. Mai sowie am 15. Dezember jeweils um 11.00

Uhr, ermitteln wir die Gewinner-Sites, die wir auch auf heise online veröffentlichen werden.

Der Wettbewerb soll einen Einblick in die Ranking-Mechanismen der Suchdienste und aktuelle Trends der Optimierung – legitime wie unerwünschte – ermöglichen. Das Vorbild Nigritude Ultramarine hat im letzten Jahr einige neue Entwicklungen ans Licht gebracht [5].

Um es den Teilnehmern nicht zu einfach zu machen, haben wir die Site www.hommingberger-gepardenforelle.de bereits angelegt, mit dem überarbeiteten Auftritt des virtuellen Fischzüchters Danuber aus dem Artikel auf Seite 158 belegt und von heise online (Page Rank 8) aus verlinkt. Unter www.heise.de/ct/SEO-Wettbewerb/ finden Sie ein Forum zum Wettbewerb.

