



Positionskampf

Web-Seiten für Suchmaschinen optimieren

Rainer Gievers/mb

Wer zuerst kommt, mahlt zuerst - dieser Grundsatz gilt auch für die Suchmaschinen, denn wessen Web-Seiten zuerst angezeigt werden, der kann eher damit rechnen, daß sie auch besucht werden. Wir zeigen Ihnen in diesem Workshop, wie Sie mit wenig Aufwand die Platzierung Ihrer Homepage in den Suchmaschinen verbessern

Die Größe des World Wide Web stellt für Suchmaschinen ein fast unüberwindliches Problem dar, denn laufend kommen neue Web-Seiten hinzu, andere werden aktualisiert oder entfernt. Zwar sind bei den Search-Engines pausenlos Crawler im Einsatz, die sich im Web von Link zu Link hangeln, trotzdem können einige Wochen oder sogar Monate vergehen, bis neue beziehungsweise aktualisierte Seiten Eingang in die Datenbanken der Suchmaschinen finden. Sie sollten sich daher nicht darauf verlassen, daß irgendwann einmal ein Crawler auf Ihrer Homepage vorbeischaud. Diese können zum einen nur Seiten finden, auf die Links von bereits katalogisierten Seiten verweisen. Zum anderen haben selbst die größten Suchmaschinen wie AltaVista und Hotbot weniger als 50 Prozent des Internets erfaßt, wenn man den Schätzungen von bis zu 240 Millionen Web-Seiten im WWW Glauben schenkt. Wir empfehlen Ihnen deshalb, nicht zu warten, bis ein Crawler Ihre Homepage erfaßt, sondern Ihre Web-Seiten

direkt bei den verschiedenen Search-Engines anzumelden. Wichtig ist auch die Anmeldung zumindest bei den größeren Internet-Katalogen wie Web.de oder Dino-Online.

Bei der Erfassung der Seiteninhalte gehen die Suchmaschinen unterschiedlich vor. Während einige nur den sichtbaren Teil der Seiten auswerten, speichern andere zusätzlich noch Teile der Meta-Tags. Weitere lesen nur einige hundert Zeilen oder - wie Crawler.de - nur eine bestimmte KB-Zahl ein. Search-Engines, die auch den Seitentext speichern, versuchen häufig, über das Auslassen bestimmter Begriffe, den Stopwörtern, Platz zu sparen. Dazu gehören AltaVista, Excite, Hotbot und Lycos. Betroffen sind davon vor allem Wörter wie „the“ oder Begriffe, die mit dem Internet zu tun haben. Optimiert man seine Seiten auf diese Suchmaschinen hin, sollte man durch Eingabe der auf der Web-Site herausgestellten Begriffe in den Suchmaschinen überprüfen lassen, ob sie überhaupt in den Datenbanken gespeichert werden.

Ranking

Nach erfolgter Suche führt jede Search-Engine ein Ranking durch, bei dem ermittelt wird, welche Web-Seiten den Suchbegriffen am nächsten kommen. Die URLs werden dann dem Ranking entsprechend ausgegeben. Um Manipulationen vorzubeugen, hält natürlich fast jeder Suchmaschinenbetreiber den genauen Ran-

Glossar

Crawler, auch als Robots oder Spiders bezeichnet, sind Programme, die Web-Sites besuchen und sich aufgrund der Hypertext-Struktur des Webs systematisch von Web-Seite zu Web-Seite bewegen. Crawler werden von Suchmaschinen eingesetzt, um das Web zu katalogisieren.

Meta Tags befinden sich am Anfang jeder Web-Seite, werden jedoch im Browser nicht angezeigt. Sie enthalten Informationen, die von den Suchmaschinen und den Internet-Proxy-Servern ausgewertet werden.

Ranking: Das von jeder Suchmaschine durchgeführte Ranking entscheidet darüber, welche URLs am ehesten dem Suchbegriff entsprechen.

Stopwörter: Um Platz zu sparen und die Abfragen zu beschleunigen, werden von den Volltext-indexierenden Suchmaschinen nur Teile der Web-Seiten gespeichert. Vermeintlich irrelevante „Stopwörter“ wie „the“ oder „web“ werden entfernt.

Suchmaschinen und **Internet-Kataloge** sind zwar beide mit einer Suchfunktion ausgestattet, unterscheiden sich aber darin, daß erstere selbstständig über Crawler-Programme die Seiten zusammensuchen, während letztere redaktionell betreut werden. Um in die Internet-Kataloge aufgenommen zu werden, müssen Sie sich bei diesen anmelden. In der Regel beträgt daher der Umfang der Internet-Kataloge nur einige hunderttausend, der Umfang von Suchmaschinen mehrere Millionen Seiten, was auch daran liegt, daß Internet-Kataloge meist nur die Startseite speichern.

Imagemaps sind Grafiken, die mehrere Links enthalten. Je nachdem, auf welchen Bereich der Anwender klickt, wird auf einen anderen Link verzweigt.

Interessante Links

URL	Titel	Beschreibung
www.suchfibel.de/	Die Suchfibel	Alles über die Kunst des Suchens
www.searchengine watch.com/	Search Engine Watch	Ausführliche Hintergrundinfos zu Suchmaschinen und deren Funktionsweise
www.deadlock.com/ promote/	Deadlock Design	Allgemeine Hinweise über die Positionsverbesserung in Suchmaschinen
www.faqs.org/faqs/ internet/info-research-faq/ part1/	Information Research FAQ	Suche im Internet

king-Mechanismus geheim, es gibt aber einige allgemeine Regeln, an die sich viele Suchmaschinen halten.

Einen großen Stellenwert beim Ranking nehmen die hervorsteckenden Merkmale einer Web-Seite ein, das heißt URL, Seitentitel, Meta Tags und Überschrift.

Während bei letzterem wohl kaum eine Änderung möglich ist, ohne die Seite unansehnlich werden zu lassen, kann der Seitentitel einfach angepaßt werden. Dabei sollten so viele auf den Seiteninhalt bezogene Begriffe wie möglich vorkommen. Auch die URL sollte im Bezug zum Seiteninhalt stehen, zum Bei-

spiel www.xyz.de/handheld/psion.htm, wenn es um ein bestimmtes Produkt geht. Eine Übersicht der Meta Tags finden Sie im Kasten 1 auf S. 109.

Wichtig ist natürlich auch die Häufigkeit der gefundenen Suchwörter, und zwar sowohl auf jede einzelne Seite bezogen als auch

im Vergleich der Seite zum Gesamtindex aller gefundenen Seiten. Teilweise wird die Relevanz einer Seite auch dadurch ermittelt, daß die Anzahl der Wörter einer Seite durch die gefundenen Suchwörter geteilt wird. Daher sollten Seiten, die sich mit vielen unterschiedlichen Themen befassen, nicht nur allein der Übersicht halber in mehrere Einzelseiten aufgesplittet werden.

Auch Faktoren, auf die Sie praktisch keinen Einfluß haben, gehen bei einigen Search-Engines ins Ranking ein. Zeigen viele Links auf Ihre Homepage, so ist dies ein Indikator dafür, daß dort relevante Informationen angeboten werden. Außerdem arbeiten inzwischen viele amerikanische Suchmaschinenbetreiber mit Re-

Suchmaschinen und Internet-Kataloge

Deutschsprachige Suchmaschinen/Internet-Kataloge

Name	Typ	URL	indexierte Seiten	Abfragen/Monat	Ranking bzw. Anzeige basiert auf
Alles Klar	K	www.allesklar.de	120.000 (D)	0,8 Mio	1. redaktionell vergebener Titel; 2. redaktionell vergebener Untertitel, Kurzbeschreibung; 3. Stichwörter; 4. Kategorie inkl. Synonyme; 5. Ort 6. Name des Anmelders; 7. URL 8. Meta Tags: Title, Description, Content
Bellnet	K	www.bellnet.com	125.000	1,6 Mio	Titel, Beschreibung, Meta Tag „keywords“
Crawler.de	S	www.crawler.de	4,8 Mio (D)	1,8 Mio	Server-Name, URL, Titel, Meta-Tags, Pfadtiefe, Sprachqualität
Dino Online	K	www.dino-online.de	170.000 (D)	6,5 Mio	URL, Beschreibung, Kategorie, Stichwörter
Eule	S	www.eule.de	2 Mio. (D)	0,936 Mio	Titel, Keywords, URL
Excite	S	www.excite.de	50 Mio	k.A.	Bewertung von selbsterstellter Textzusammenfassung, Link Popularität
Fireball	S	www.fireball.de	7 Mio (D)	25 Mio.	Express-Suche: erstes Auftreten des Suchworts in Seite, Häufigkeit des Suchworts, bei mehreren Suchwörtern: Distanz zwischen den Begriffen. Allgemein: Seitentitel, URL, Überschriften, Häufigkeit des Suchworts im Text, Häufigkeit des Suchworts im Gesamtindex, Hervorhebungen, Meta Tag „keywords“, Datum.
Infoseek Germany	S	www.infoseek.com/ Home?pg=Home.html&sv=DE	30 Mio	k.A.	Anzahl der Suchwörter in Seite und im Vergleich zu allen anderen gefunden
Lycos Deutschland	S	www.lycos.de	100 Mio.	20 Mio	Volltext, Meta Tags
Web.de	K	www.web.de	161.000 (D)	15 Mio	Internes Punkteschema: Suchwort in: URL: 5 Punkte; Namen: 5 Punkte; Wortgrenze: zusätzlich 1 Punkt; am Beginn des Namens: zusätzlich 2 Punkte. Jeder Treffer im Infoteil: 1 Punkt; auf einer Wortgrenze: zusätzlich 1 Punkt; am Beginn des Infoteils: zusätzlich 1 Punkt. Bei Mehrwortsuche erfolgt eine Addierung
Yahoo Deutschland	K	www.yahoo.de	80.000 (D)	k.A.	URL, Titel, Beschreibung, Kategorie

view-Services zusammen, die ausgewählte Web-Sites bewerten. Findet die Suchmaschine solche Seiten, wird deren Ranking ebenfalls erhöht.

Beachten Sie, daß die Search-Engines unterschiedlichen Konzepten folgen, so daß eine Optimierung, die zu einer Verbesserung der Position in einer Suchmaschine führt, bei anderen wiederum einem Positionsrückgang zur Folge haben kann.

Probleme beim Indexieren

Schon in der Design-Phase sollten Sie einige Dinge beachten, um den Suchmaschinen die Indexierung Ihrer Seiten zu erleichtern. So können viele Suchmaschinen mit dynamisch - also mit CGI-BIN-Skripten - erzeugten

Web-Seiten nichts anfangen und werden diese nicht auswerten. Das kann bedeuten, daß statt hundert dynamisch erzeugter Seiten nur die Hauptseite, die diese Seiten erzeugt, im Katalog gespeichert wird. Entsprechend gering ist dann die Wahrscheinlichkeit, daß die Seiten gefunden werden. Ändern sich die Inhalte der dynamisch erzeugten Seiten nur gelegentlich, sollten Sie daher das Skript, das die dynamischen Seiten erzeugt, so abändern, daß es diese als normale HTML-Seiten-Dateien anlegt.

Ein weiteres Problem werfen Web-Sites auf, die in Frame-Technik ausgeführt sind. Viele Crawler, zum Beispiel der von Excite, können Frames nicht verarbeiten, und ausgerechnet auf der Startseite (wo die Frames zusammengeführt werden) ist meist kein indexierbarer Text vorhanden, mit der Folge, daß die komplette Web-Site nicht katalogisiert wird. Es empfiehlt sich daher, auf der Startseite im Abschnitt `</noframes>` zumindest eine kurze Beschreibung unterzubringen. Ist eine Frame-basierte Homepage einmal katalogisiert, besteht die nächste Schwierigkeit darin, daß Suchmaschinen nur URLs auf Teil-Frames anzeigen. Enthalten diese keine Navigationselemente, kommt der Nutzer von dort nicht weiter zu den anderen Seiten der Web-Site. Daher sollten Sie auf jeder Seite zumindest einen weiterführenden Link (zum Beispiel auf die Startseite) anbringen. Möchten Sie über diesen Link auch gleich veranlassen, daß die Frames neu aufgebaut werden, sollten Sie ihn auf die Frame-Hauptseite mit dem `_top`-Parameter verweisen lassen. Beispiel: ``.

Ähnliche Schwierigkeiten wie Frames bereiten einigen Crawlern auch Imagemaps. Weil auch manche Browser keine Imagemaps unterstützen, sollte auf jeder mit Imagemaps ausgestatteten Web-Seite zusätzlich die Navigation über Textlinks möglich sein.

Aus verschiedenen Gründen, zum Beispiel wenn die Bandbreite ei-

Meta Tags

Viele Meta Tags haben nur informativen Charakter und werden von den Suchmaschinen beim Ranking nicht ausgewertet, teilweise aber bei der Suchergebnisanzeige mit ausgeworfen. Es lohnt sich also, bei „description“ eine kernige und genaue Beschreibung des Seiteninhalts zu geben.

```
<meta name=„author“ content=„Rainer Gievers“>
<meta name=„copyright“ content=„PC INTERN“>
<meta name=„description“ content=„PC INTERN, das Fachmagazin für PC-Profis“>
```

Von größerer Bedeutung ist das Meta Tag „keywords“, weil dieses bei vielen (aber beileibe nicht allen) Suchmaschinen für das Ranking genutzt wird. Fallen Ihnen keine passenden Begriffe ein, schauen Sie sich einfach einmal die Meta Tags von Konkurrenz-Web-Seiten an. Auch wenn die Länge der Tags nicht durch irgendwelche Standards vorgeschrieben ist, sollten Sie beachten, daß Suchmaschinen Meta Tags nur bis zu einer bestimmten Länge auswerten. Stellen Sie daher die wichtigeren Begriffe nach vorne. Bezüglich der Verwendung des Plurals bei den Wörtern können wir keine allgemeine Empfehlung abgeben, es dürfte aber zum Beispiel eher nach „Downloads“ statt nach „Download“ gesucht werden.

```
<meta name=„keywords“ content=„Notebook, Computer, Windows, Magazin, News, Meldungen, kostenlos, Kleinanzeige, Shareware, Hardware, Freeware, Kaufberatung“>
```

Dieser Meta Tag soll darauf hinweisen, daß sich die Seite regelmäßig ändert und sie daher häufiger von den Crawlern besucht werden soll. Eine Garantie dafür gibt es allerdings nicht.

```
<meta name=„revisit-after“ content=„2 days“>
```

Zur Steuerung der Crawler steht das Meta Tag „robots“ zur Verfügung. Im ersten Beispiel werden Crawler angewiesen, die Seiten zu indexieren und den Links zu folgen, im zweiten wird beides verboten.

```
<meta name=„robots“ content=„index, follow“>
<meta name=„robots“ content=„noindex, nofollow“>
```

nen Servers begrenzt ist, kann es auch einmal notwendig sein, Seiten von der Indexierung durch die Crawler auszuschließen. Dafür wurde 1994 das Robot Exclusion Protocol geschaffen, das in einer einfachen Textdatei „robots.txt“ angibt, welche Verzeichnisse einer Web-Site nicht von den Crawlern besucht und indexiert werden sollen. Die meisten Anwender, wozu alle Privatnutzer von Online- und Internet-Zugangsdienstleistern zählen, haben allerdings keinen Zugriff auf das Root-Verzeichnis. Als Ersatz bietet sich zum Glück der Meta Tag „robots“ an. Betreiben Sie ei-

ne eigene Web-Site mit Seiten, die nicht indexiert werden sollen, empfehlen wir auf jeden Fall einmal die Dokumentation zum Robot Exclusion Protocol zu studieren, denn dieses ist wesentlich praktischer zu handhaben (Sperren von ganzen Verzeichnissen mit einem Befehl) als der „robots“-Meta Tag und bietet zusätzliche Optionen.

Was kommt danach

Von der Anmeldung Ihrer Seiten bei einer Suchmaschine bis zur Aufnahme in den Index können ein Tag bis zu acht Wochen vergehen. Tauchen die Seiten dann

Gespeichert werden
Titel, Stichwörter, Anmeldename, MD (alle Meta Tags zusammen max. 2.500 Zeichen)
Titel (60), Beschreibung (200), MK (100) V (max. 64 KB), KC
Beschreibung, Stichwörter MK (255), MD (255), Titel (60), KC KF, VS
V, MK (256), MD (256)
V, KF, MD (200), MK (1000)
VS, MK, MD
Titel, Beschreibung
Titel, Beschreibung



Fireball unterhält unter www.fireball.de/metagenerator.html eine Seite, die Sie bei der Erzeugung korrekter Meta Tags unterstützt

Englischsprachige Suchmaschinen/Internet-Kataloge

Name	Typ	URL	Anzahl indizierter Seiten	Abfragen/Monat	Ranking basiert auf	Gespeichert werden
AltaVista	S	www.altavista.com	140 Mio.	k.A.	Seitentitel, Seitentext, höhere Bewertung von Worten am Anfang. Nähe der gesuchten Wörter zueinander	VS
Excite	S	www.excite.com	55 Mio.	1,5 Mrd.	Bewertung von selbsterstellter Textzusammenfassung, Link-Popularität	KF, V, KI
Hotbot	S	www.hotbot.com	110 Mio.	k.A.	Titel, Meta Tags, Anzahl der Suchwörter auf Seite geteilt durch Wörteranzahl auf Seite	VS, MK, KI
Infoseek	S	www.infoseek.com	30 Mio.	600 Mio.	Anzahl der Suchwörter in Seite und im Vergleich zu allen anderen gefundenen Seiten, Link-Popularität, Meta Tags	V, KF, MD (200), MK (1000)
Lycos	S	www.lycos.com	30 Mio.	1,5 Mrd.	Volltext, Meta Tags, Link-Popularität	VS, MK, KI
Webcrawler	S	www.webcrawler.com	2 Mio.	k.A.	Anzahl der Suchwörter auf Seite geteilt durch Wörteranzahl auf Seite, Link-Popularität	VS, KF
Yahoo	K	www.yahoo.com	1 Mio.	k.A.	URL, Titel, Beschreibung, Kategorie	Titel, Beschreibung

V = speichert Volltext; VS = speichert Volltext ohne Stopwörter; MK = Auswertung des Meta Tags „keywords“; MD = Auswertung des Meta Tags „description“; KF = keine Frames; KC = es werden keine durch CGI-BIN-Skripte erzeugte Seiten aufgenommen; KI = keine Imagemaps; (D) = nur deutschsprachige Seiten; Zahlen in Klammern geben an, wie viele Zeichen jeweils ausgewertet werden.

immer noch nicht auf, sollten Sie die Anmeldung einfach wiederholen. Haben Sie später einmal Änderungen an Ihren Seiten vorgenommen, beschleunigen Sie das Update der Datenbank durch eine erneute Anmeldung.

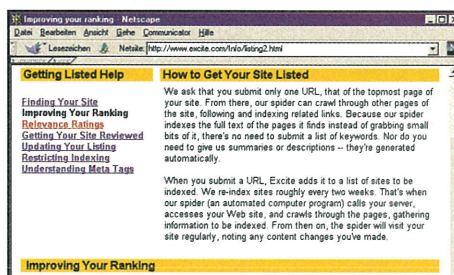
Tricks, die danebengehen

Immer wieder wird versucht, das Ranking in den Suchmaschinen durch Tricks zu verbessern - besonders „Schmuddel-Sites“ scheinen dafür anfällig zu sein. Sehr beliebt ist es zum Beispiel, bestimmte Wörter endlos am Seitenende beziehungsweise im Meta Tag „keywords“ zu wiederholen oder unsichtbare Texte (Farbe wie Hintergrund) unterzubringen. Auch Texte in sehr kleiner Schrift werden von einigen Search-Engines unter Umständen als Versuch der Ranking-Beeinflussung wahrgenommen. Die Suchmaschinen reagieren auf diese „Spamming“ genannten Tricks höchst unterschiedlich: Einige

nehmen solche Seiten überhaupt nicht in den Index auf, andere reduzieren zur Strafe das Ranking, weitere lassen verdächtige Seitenbestandteile beim Ranking einfach außen vor. Großer Popularität erfreut sich auch das Erstellen von möglichst vielen gleichartigen Seiten nach dem Motto „viel hilft viel“. Einige Suchmaschinen erstellen allerdings eine Prüfsumme aller indizierten Seiten und erkennen so Doppelgänger, die entweder aus dem Index gelöscht oder bei der Suchergebnisausgabe zu einem Eintrag zusammengefaßt werden. Wegen der unvorhersehbaren Folgen raten wir vom Spamming ab.

WebPosition

Die Kontrolle der eigenen Platzierung in den verschiedenen Suchmaschinen kann ganz schön in Arbeit ausarten, vor allem, wenn man mehrere verschiedene Suchbegriffskombinationen ausprobieren muß. WebPosition von 1st



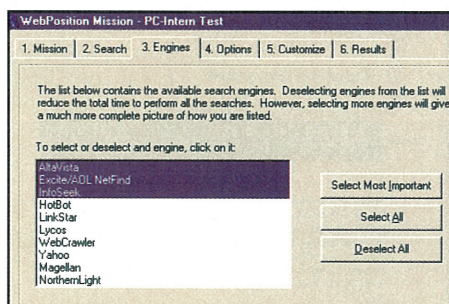
Fast jede Search-Engine informiert ausführlich über mögliche Suchmethoden, tieferegehende Infos über die Ranking-Mechanismen sucht man aber meist vergebens

Source übernimmt und automatisiert diesen Vorgang für die zehn wichtigsten Suchmaschinen und Internet-Kataloge. In der Trial-Version, die Sie auch auf der PC-INTERN-Heft-CD finden, werden AltaVista, Excite und Infoseek unterstützt. Nach Eingabe der zu beobachtenden Web-Seiten und der Suchwortliste führt Web-Position automatisch die Suchanfragen durch. Das Ergebnis wird dann schließlich in HTML-Dateien gespeichert, die im Web-Browser angezeigt werden können. Bei unseren ersten Tests konnte Web-Position allerdings nichts mit den Ausgaben von AltaVista anfangen, da sich offenbar das Ausgabeformat geändert hatte. Nachdem wir die komfortable Internet-Update-Funktion ausgeführt hatten, funktionierte die Software aber schließlich einwandfrei. In den generierten Reports wird unter anderem festgehalten, welche Position die eigenen Web-Seiten in den gewählten Suchmaschinen haben; eine History zeigt an, ob

es Positionsverschiebungen gegeben hat. Um Telefongebühren zu sparen (bei einem Dutzend Suchwortkombinationen kann es schon einmal eine Viertelstunde werden), ist über ein Zusatzprogramm der zeitgesteuerte Aufruf - inklusive Einwahl und Abwahlsteuerbar.

FAZIT

Für Anwender, die ihre Platzierung in amerikanischen Suchmaschinen verbessern oder verfolgen möchten, ist WebPosition sehr interessant. Die Online-Hilfe des Programms macht zudem auf eine leider nur zu häufig vergessene Binsenweisheit aufmerksam: Ohne interessante und gut aufbereitete Inhalte, die Web-Surfer immer wieder zum Hereinschauen veranlassen, ist der Einsatz von Hilfsmitteln zur Positionsverbesserung nur von begrenzter Wirkung.



Die Software Web-Position hilft bei der Feststellung der eigenen Position in verschiedenen Suchmaschinen. In der Trial-Version sind AltaVista, Excite und Infoseek verfügbar