

Semesterarbeit

Optimizing the Ranking in Google

Guido Frerker

gfrerker@student.ethz.ch

Departement Informatik

Eidgenössische Technische Hochschule Zürich

2003/2004

Prof. Dr. Roger Wattenhofer
Distributed Computing Group
Betreuer: Keno Albrecht

1	Einführung	3
2	Google-Ranking	3
3	Optimierungen	4
3.1	Schlüsselwörter im Text	4
3.2	Links	5
3.2.1	Interne Links – „Cross links“	5
3.2.2	Externe Links	5
3.3	Versteckter Text.....	5
4	Vorgehen und Analyse.....	6
4.1	Aufbau der Webseite.....	6
4.2	Optimierungen	6
4.3	Erster Google-Besuch	7
4.4	Versuche, die Webseite weiter zu optimieren.....	7
4.5	Linkfarm	8
4.5.1	Vorgehen.....	8
4.5.2	Resultat	8
5	Schlussbemerkung	9
6	Glossar:	10
7	Referenzen:	12

1 Einführung

In dieser Semesterarbeit geht es darum, eine Webseite so zu optimieren, dass sie bei Google ein möglichst gutes Ranking erhält. Dazu müssen ein oder mehrere Stichwörter ausgesucht werden, auf welche die Webseite optimiert werden soll. Diverse Tipps und Tricks, wie man eine Webseite optimieren kann, und viele Forums-Diskussionen, wie der Suchmaschinen-Algorithmus funktionieren könnte, findet man auf den WWW-Seiten von Suchmaschinen Optimierer. Eine der bekanntesten Webseite ist „Search Engine World“ [1].

2 Google-Ranking

Die „Search Engine Result Position“ (SERP) [s. Glossar], d.h. die Position an der man bei der Suchmaschine mit den ausgewählten Suchbegriffen aufgelistet wird, berechnet Google mit einem geheimen Algorithmus. Das er geheim ist sollte klar sein, denn sonst wäre das Suchmaschinenoptimieren viel zu leicht und Google somit voll von Spam [s. Glossar].

Eines der wichtigsten Punkte zur Berechnung der SERP bei Google ist der PageRank [5], [6], [7]. Allgemein auch als „Link Popularity“ bezeichnet, wird der PageRank mit der Anzahl von eingehenden und ausgehenden Links berechnet. Dabei zählen Links von bedeutenden Websites - also Sites, auf die wiederum oft verlinkt wird und deshalb auch einen höheren PageRank haben - mehr als von kleinen, unbekannt Seiten. Der PageRank wird auf einer Skala von Null bis Zehn gemessen. Zehn stellt dabei den besten Wert dar und wird z.B. vom Google-Directory erreicht. Der PageRank wird bei den Suchergebnissen leider nicht angezeigt, kann aber ermittelt werden, wenn man die Google-Toolbar [4] beim Internet Explorer installiert.

Neben dem Zählen der Links kommt es auch darauf an, mit welchen Worten die Page verlinkt wurde, dem Link Kontext. So ist es möglich, dass eine Webseite unter Suchbegriffen aufgelistet wird, die überhaupt nicht auf der Seite vorhanden sind, wie es z.B. bei der Suche nach „miserable failure“ der Fall ist (SERP 1). Das ganze ist auch als „Google bombing“ [s. Glossar] bekannt.

Neben den externen Punkten kommt es natürlich auch auf die internen, altbekannten Werte an. Dazu gehört, wie häufig ein Such- oder Schlüsselwort auf einer Seite vorhanden ist und an welchen Orten es sich befindet. Mögliche Orte sind unter anderem die URL, das „titel“ Element, die Link Texte (hyperlinked text), die Metadaten, der Anfang des Textes und innerhalb „kursiv“, „fett“ und „strong“ Tags.

Bis es aber überhaupt zu einem Ranking der Webseite kommt, kann es dauern. Mit etwa 2-3 Monaten muss man schon rechnen, bis Google die ganze Webseite indexiert hat.

3 Optimierungen

Tipps und Tricks wie man seine Webseite optimieren kann.
Zusammengetragen aus diversen Webseiten: [1], [6], [7]

3.1 Schlüsselwörter im Text

Schlüsselwortdichte:

Die Schlüsselwortdichte ist der Prozentsatz indexierbarer Wörter einer Seite und sollte so zwischen 5 und 20% betragen. Ist die Dichte zu gross kann Google die Seite unter Umständen als Spam betrachten und man wird schlecht bewertet.

Für die Berechnung der Dichte kann man z.B. [3] verwenden.

Anzahl Seiten:

Je mehr Seiten desto besser. Eine gute Anzahl am Anfang sind etwa 100 Seiten. Danach sollte man jeden Tag mindestens eine neue Seite online stellen, bei kleineren Webseiten sicher aber einmal pro Woche. Seiten mit neuem Inhalt werden von Google besser bewertet als alte, seit langem unveränderte Seiten.

Die wichtigsten Stellen in einer HTML Datei, an denen Schlüsselbegriffe stehen sollten.

Domain:

Sofern möglich, sollten die Schlüsselwörter im Domainnamen vorkommen z.B. www.schlüsselwort.com

Achtung: Google interpretiert einen Unterstrich (_) nicht als Wortabgrenzer, aber den Bindestrich (-) schon.

Dateinamen:

Anstatt die Dateien blablabla.html zu nennen, sollten sie eher schlüsselwort-irgendwas.html heissen. Auch hier sollte der Bindestrich (-) wieder zum Trennen von Worten verwendet werden.

Verzeichnisnamen:

Auch hier sollten Schlüsselbegriffe vorkommen. In jedem Verzeichnisnamen sollte mindestens ein Schlüsselwort vorkommen. Bindestrich (-) fürs Worttrennen.

Titel der Webseite („title“ Element):

Der Titel der Webseite ist eine der wichtigsten Orte, wo Schlüsselbegriffe platziert werden müssen. Aber man sollte darauf achten, dass der Titel nicht allzu lang wird, je kürzer desto besser, und für jede Seite einzigartig ist.

Metadata:

Im „Meta Description“ Tag und „Meta Keywords“ Tag sollten ebenfalls die Schlüsselbegriffe je einmal vorkommen.

Überschrift („h1“ Tag):

Die Überschrift sollte gerade nach dem „body“ Tag kommen, d.h. am Anfang der Seite. Da Google sich eine externe CSS Datei nicht ansieht, kann man die Überschrift problemlos umformatieren. Wahrscheinlich funktioniert das auch mit dem „style“ Attribut.

Formatierungs-Tags:

Die Schlüsselbegriffe sollten ebenfalls mindestens je einmal in den Folgenden Tags vorkommen:

„strong“, „i“ (*kursiv*), „b“ (**fett**)

das wichtigste davon ist das „strong“ Tag

„alt“ Attribute im „img“-Tag:

Da Google den Inhalt von Bildern (noch) nicht lesen kann, sollte man das „alt“-Attribut verwenden, um das Bild zu beschreiben. Zudem kann man hier sehr gut Schlüsselbegriffe platzieren.

3.2 Links

3.2.1 Interne Links – „Cross links“

Beim Verlinken der Webseiten untereinander sollte man vor allem Textlinks verwenden, weil Google die Bilder nicht lesen kann. Wenn man doch mit Bilder verlinken will, sollt man in jedem Fall das „alt“-Attribut verwenden.

Je mehr Seiten man hat, desto höher kann der PageRank sein. Zudem sollten alle Unterseiten auf die Startseite verweisen [6].

3.2.2 Externe Links

Man sollte versuchen, möglichst viele externe Links auf seine Seite zu bekommen, d.h. externe Seiten, die auf die eigene Webseite verlinken.

Einen enormen Schub für das Ranking in Google bringt ein passender Eintrag in Yahoo und dem Open Directory Project (ODP). Das Google-Directory ist übrigens eine Eins zu Eins Kopie des ODP.

Im Weiteren kann man Links auf seine Webseite in Gästebüchern und Foren Themenverwandter Webseiten platzieren.

3.3 Versteckter Text

Die bis jetzt aufgeführten Optimierungen sind alle „legal“, d.h. von Google akzeptiert und sollten sicher befolgt werden. Man kann aber bei der Schlüsselwortdichte etwas nachhelfen, ohne das guter, sinnvoller Text eingefügt werden muss. Mit der Hilfe von CSS Formatierungen - ausgelagert in eine CSS Datei, die Google nicht beachtet - kann man Text sehr gut verstecken, sodass er für den Besucher der Webseite nicht sichtbar ist (Schriftfarbe = Hintergrundfarbe). Weil das verstecken von Text aber von Google nicht geduldet wird, kann es sein, dass die Webseite ein Strafe erhält und somit eine schlechtere Rangierung hat oder gar ganz aus Google rausgeworfen wird. Das kann aber nur passieren, wenn der versteckte Text von jemandem entdeckt und die Webseite dann bei Google angezeigt wurde.

Das verstecken des Textes sollte man nicht mit dem „Cloaking“ [s. Glossar] verwechseln, wo einem Surfer eine andere Seite gesendet wird als der Suchmaschine.

4 Vorgehen und Analyse

4.1 Aufbau der Webseite

Das Thema der Webseite lautete Kinofilme. Als Suchbegriffe wurden „Kino-Film“, „Kino“ und „Film“ ausgewählt, wobei der Schwerpunkt auf „Kino-Film“ gelegt wurde. Dabei wurde nicht gross darauf geachtet, mit welchen Begriffen Google Benutzer am häufigsten suchen, sondern eher dass auch ein Resultat erhofft werden durfte. Gesucht wird mit:

<http://www.google.ch/ie?q=kino-film&num=100&cr=countryCH>

Damit schnell sinnvoller Inhalt eingefügt werden konnte, wurde der meiste Text mit freundlicher Erlaubnis von OutNow.CH [2] kopiert, eine gut besuchte Webseite rund ums Kino.

Die Webseite wurde so aufgebaut, dass von der Hauptseite auf die 6 aktuellsten Filme (mit je 2 Links) und eine Archiv-Seite verlinkt wird. Pro Film gibt es genau eine Seite und jeder Film ist auf der Archiv-Seite aufgeführt. Alle Seiten haben einen Link auf die Startseite und alle Filmseiten zusätzlich einen Link auf die Archiv Seite. Jede Seite ist deshalb mit maximum 2 Klicks erreichbar.

Die URL der Webseite lautet www.kino-film.ch. Filme sind im Verzeichnis „kino-film/filmname“ abgelegt.

4.2 Optimierungen

Ziel der Optimierungen ist, dass vor allem die Startseite eine gute Rangierung erhält. Dazu wurde das Schlüsselwort „Kino-Film“ wie folgt verteilt:

- Der Titel („title“ Element) enthält genau das Schlüsselwort.
- Die Überschrift („h1“ Tag) besteht auch aus genau dem Schlüsselwort.
- Im Text kommt das Schlüsselwort einmal in einem „strong“ Tag und dreimal normal vor.
- Metadaten: „Kino-Film“, „Kino“ und „Film“ in Meta Keywords und „Kino-Film“ in Meta Description.
- „alt“ Attribut: Die 6 Filme auf der Startseite werden mit je 2 Links verknüpft, einem Textlink „ Filmname“ und einem Bild. Dabei steht im „alt“ Attribut des Bildes: „Kino-Film: Filmname“.

Daraus ergibt sich eine Schlüsselwortdichte von etwa 12% (berechnet mit [3]).

Die so erstellte Webseite wird nun bei Google [8] angemeldet und bei den Foren von [2] und [9] ein Link eingetragen. Einen Eintrag bei Yahoo und dem ODP gab es leider nicht, weil die Seite mit nur kopiertem Inhalt den Ansprüchen der jeweiligen Verzeichnisse nicht entsprach.

Damit die Webseite auf einem aktuellen Stand ist, wenn Google das erste Mal vorbeikommt, muss sie regelmässig erweitert oder zumindest abgeändert werden.

4.3 Erster Google-Besuch

Das erste Mal kam Google nach ca. einem Monat vorbei, schaute sich aber nur die Startseite an. Trotzdem wird die Startseite auf dem 322 Platz in der Suche aufgelistet. Einen PageRank gibt es noch nicht. [s. Diagramm]

Nach zwei weiteren Besuchen der Startseite, ca. 4 bzw. 7 Wochen später, rutschte der Rang zuerst auf Platz 400 ab und sprang dann auf Platz 100 vor. Der Sprung ist vermutlich auf das Update im Februar 2004 des Google Algorithmus zurückzuführen („Brandy“-Update). Dabei kommt es vor, dass der SERP der Webseite auf und ab hüpfte → 'dancing' Google [s. Glossar].

Zwei Monate nach dem ersten Googlebesuch wird die gesamte Webseite besucht. Zu diesem Zeitpunkt besteht die Webseite aus 23 Filmseiten. Eine Woche danach wird das erste Mal der PageRank ausgerechnet. Die Startseite erhält den PR 3 und bei der Suche landet sie auf der Position 51.

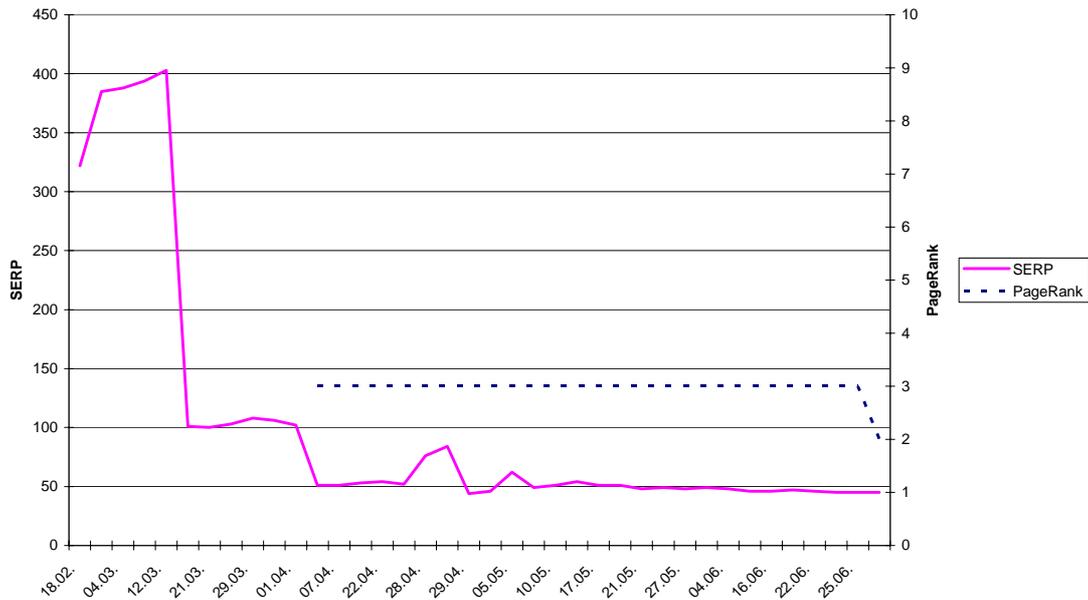
4.4 Versuche, die Webseite weiter zu optimieren

Nach der ersten vollständigen Indexierung der Webseite kommt der Googlebot regelmässig ca. alle 12 Tage vorbei. In dieser Zeit sind folgende Änderungen gemacht worden [s. Diagramm]:

- {08.04.} ca. 50 Begriffe, die alle mit Kino und Filme in Beziehung stehen, wurden als versteckter Text am Ende der Seite hinzugefügt, wodurch sich die Schlüsselwortdichte auf 11% senkt
- {08.04.} alle Links mit der absoluten Pfadangabe (vorher relative Pfadangabe)
- {20.04.} alle Filmseiten werden von „index.html“ in „kino-film.html“ umbenannt
- {22.04.} alle Links wieder mit relativer Pfadangabe
- {22.04.} das Archiv von 30 auf etwa 70 Filme (= 70 Seiten) erhöht
- {12.05.} auf allen Seiten zwei externe Links auf Themenverwandte Seiten eingefügt (OutNow.CH und IMDB.com)
- {06.06.} das Archiv in mehreren Schritten auf etwa 90 Filme (= 90 Seiten) erhöht

Alle diese Veränderung haben aber keine sichtbaren Veränderungen zur Folge und die Platzierung bleibt immer um den 50. Rang herum. Eine Ausnahme bleibt einmal der Rutsch auf Rang 80, was vermutlich auf die Umbenennung der Filmdateien zurückzuführen ist. Damit alte Links nicht ins Leere führen, sind die Filme für einen Google-Besuch doppelt vorhanden (als „index.html“ und als „kino-film.html“). Google hat sich nun beide Dateien angeschaut und dabei natürlich Duplikate erkannt (z.B. gleicher Titel und Dateigrösse), was eventuell zu Strafpunkten geführt hat.

Verlauf der Rangierung



4.5 Linkfarm

Als nächstes sollte getestet werden, ob man mit zusätzlichen kostenlosen Domains, mit den Endungen wie „.ch.vu“, „.ch.tf“, „.ch.tt“, die auf die eigene Webseite mit Frames verlinken, gratis externe Links erhalten kann.

4.5.1 Vorgehen

Damit ein Effekt erwartet werden konnte, wurden ca. 70 Domains registriert und alle mit der „kino-film.ch“ Webseite verknüpft.

Damit Google die links schneller in den Index aufnehmen kann, wurde bei einem „kostenlos Webspace“ Anbieter (z.B. Geocities) eine Webseite eingerichtet, die auf alle Gratisdomains einen Link hat.

4.5.2 Resultat

Weil Google neue Seiten leider nur sehr langsam in seine Datenbank aufnimmt (mehr als zwei Monate Wartezeit), lag ein Ergebnis bis zum Fertigstellen dieser Arbeit leider nicht vor.

5 Schlussbemerkung

Das Optimieren von Webseiten ist ein sehr langsamer Vorgang und erfordert deshalb auch eine über eine weite Zeitspanne verteilte Arbeit. Vor allem zu Beginn, bis man seine Webseite vollständig in Google drin hat, dauert es Monate.

Meine Semesterarbeit erstreckte sich deshalb über einen Zeitraum von etwa einem halben Jahr. Während dieser Zeit hat Google den Algorithmus zum Berechnen des Ranges zweimal geändert, was in den SEO-Foren [s. Glossar] immer wieder zu interessanten, bücherfüllenden Diskussionen geführt hat. Auch wurden immer wieder neue Theorien aufgestellt, die aber oft widerlegt wurden. Nur wenige, wie der „Sandbox“ Effekt [s. Glossar], konnten sich halten.

Mit den in diesem Dokument aufgeführten Optimierungen kann man erstaunlicherweise schon eine recht gute Rangierung bei Google erreichen, auch wenn man nur sehr wenige Backlinks auf seine Seite hat. Und wenn man auf versteckten Text und auf Linkfarmen verzichtet, kann auch niemand etwas dagegen haben.

6 Glossar:

„Cloaking“:

Dabei wird der Suchmaschine eine andere Version der Webseite zurückgegeben als einem normalen Benutzer. Durch die IP-Adresse oder dem User-Agent (z.B. „Googlebot/2.1 (+http://www.googlebot.com/bot.html)“) kann man die Suchroboter von normalen Besuchern unterscheiden.

'dancing' Google:

Die SERP einer Webseite Hüpf auf und ab.

Tritt meist bei einem Update der Suchmaschine auf, d.h. wenn Google den Algorithmus geändert hat, weil die einzelnen Datenzentren nacheinander aktualisiert werden. Diesen Vorgang kann man z.B. mit [10] beobachten.

Das letzte Google-Update mit dem Namen „Brandy“ wurde Mitte Februar durchgeführt.

<http://www.webmasterworld.com/forum3/24435.htm>

Google bombing:

Auf einer grossen Anzahl von Webseiten Links einfügen, die alle auf dieselbe Webseite mit dem gleichen Linktext zeigen, sodass die Seite zuoberst bei einer Suche erscheinen, wenn man den Linktext eingibt.

Bsp: „miserable failure“

<http://uber.nu/2001/04/06/>

PageRank:

Ähnlich wie das Konzept der Link-Popularität, ist die absolute Anzahl eingehender Links für die Beurteilung der Bedeutung einer Webseite von Nutzen. Zudem werden aber die Links unterschiedlich gewichtet.

<http://www.newscriptor.com/clixtrax/glossar/definition/PageRank.html>

Sandbox Effekt:

Neue Seiten werden nur auf den unteren Rängen platziert, obwohl sie viele gute Links von Aussen haben. Erst wenn die Seite ein gewisses „Alter“ erreicht (ca. 2-4 Monate), wird sie aus dem Sandkasten entlassen und kann beim normalen Ranking mitspielen. In den Sandkasten platziert wird man, indem Google die Backlinks – die externen Links auf eine Seite – nicht berücksichtigt, was zu einem schlechten oder überhaupt keinen PageRank führt.

Es wird vermutet, dass man damit das PageRank Erwerben durch Google bombing verhindern will

<http://www.seroundtable.com/archives/000368.html>

<http://www.webmasterforum.ch>

SE:

Search Engine

SEO:

Search Engine Optimization

SERP:

Search Engine Result Position

Numerischer Rang, der die Position in einer Resultatliste reflektiert, wenn jemand eine Suche mit einem bestimmten Satz von Schlüsselwörter startet.

Spam (bei Suchmaschinen):

Suchergebnisse, die nichts mit der Suchanfrage zu tun haben.

<http://www.spiegel.de/spiegel/0,1518,272384,00.html>

7 Referenzen:

- [1] Search Engine World By Webmasters For Webmasters
<http://www.searchengineworld.com/>
<http://www.webmasterworld.com/>
<http://www.searchengineworld.com/misc/guide.htm>
http://www.searchengineworld.com/spiders/google_faq.htm
<http://www.webmasterworld.com/forum5/3507.htm>
- [2] OutNow.CH - die andere Seite des Films...
<http://www.outnow.ch/>
- [3] Keyword Density Analyzer
<http://www.keyworddensity.com>
<http://www.ranks.nl/tools/spider.html>
- [4] Google Toolbar
<http://toolbar.google.com>
- [5] Suchmaschine Google | Wie funktioniert der Mini-Page-Rank?
http://www.rankhigher.de/minirank_info.htm
- [6] Pagerank Explained. Google's PageRank and how to make the most of it.
<http://webworkshop.net/pagerank.html>
- [7] Suchmaschinen Tricks: Suchmaschineneintrag, Suchmaschinen Optimierung und Suchmaschinen Anmeldung
<http://www.suchmaschinentricks.de/>
- [8] URL bei Google hinzufügen
<http://www.google.ch/intl/de/addurl.html>
- [9] Das Kino-Forum
<http://www.kino-forum.com>
- [10] My Google - Dance – Machine
<http://google-dance.miniunternehmen.de>

weiter Seiten über SEO

<http://www.searchguild.com>
<http://www.highrankings.com>
<http://www.cre8asiteforums.com/index.php>
<http://www.internet-marketing-research.net>
<http://www.jimworld.com>
<http://www.seobook.com>
<http://www.abakus-internet-marketing.de/index.shtml>
http://www.portentinteractive.com/services_seo.htm

Webseiten-Verzeichnisse

<http://www.dmoz.de>
<http://www.yahoo.de>
<http://directory.google.com>
<http://www.allesklar.de>
<http://web.de>

gratis Domainname Anbieter:

<http://www.nic.ch.vu/>
<http://www.joynic.com/>
<http://www.alpennic.com/>

alles über Google

<http://www.google.ch>

<http://googlefan.com/modules/Everything>

<http://www-db.stanford.edu/~backrub/google.html>

<http://www.at-web.de/google/>