

ARD.ZDF
medien
akademie

Henk van Ess

Der Google-Code

Das Geheimnis der besten Suchergebnisse

GOOGLE



ADDISON-WESLEY

In diesem Kapitel lernen Sie eine eigenwillige, jedoch äußerst effektiven Suchmethode kennen. Richten Sie keine Fragen mehr an Google, sondern nehmen Sie den Anfang einer Antwort vorweg. Sie lernen außerdem, dass Parkinson nicht tödlich ist.

Google ist kein Orakel, sondern eine Maschine. Sie kennt den Inhalt von Milliarden von Webseiten und vergleicht diesen mit Ihrer Suchanfrage. Wenn Sie in Google suchen, dann spielen Sie gewissermaßen eine Partie Memory.



Wenn Sie Ihre Frage auf den Tisch legen, sucht Google die dazugehörige Antwort. Die Maschine nimmt Ihre Frage sehr wörtlich.

Die Intelligenz muss von Ihnen kommen, nicht von Google. Verlassen Sie sich zu sehr auf die Suchmaschine, geht tagtäglich sehr viel schief. Ihre Ergebnisliste wird zu lang, zu unklar, zu kurz, unzuverlässig, undeutlich, nicht interessant, unlesbar oder unsinnig. Schon ein kleiner Buchstabe kann den Unterschied ausmachen sein. Nehmen wir z.B. an, Sie suchen mit der folgenden Anfrage nach einem Foto von einem Fiat Punto:

Google

Suche

Ungefähr 39.800.000 Ergebnisse (0,19 Sekunden)
Erweiterte Suche

Alles
 Bilder
 Mehr

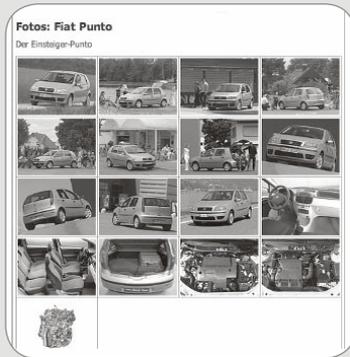
Fotopunto: [modelos](#), [fotógrafos](#), [agencias de modelos](#), [maquilladoras](#)
... - [Diese Seite übersetzen]
Date de alta como modelo, fotógrafo, maquilladora, estilista, agencia de modelos , etcétera
... Selección y casting de modelos y profesionales de la ...
www.fotopunto.com/ - Im Cache - Ähnliche

Das Web
Modelos
Zona de gestión

Stattdessen landen Sie bei einer spanischen Fotomodell-Site. Mit *Foto Punto* finden Sie als erstes Ergebnis nichts Brauchbares, mit *Fotos vom Punto* aber schon:

The screenshot shows a Google search interface. The search bar contains the text "Fotos vom Punto". Below the search bar, it indicates "Ungefähr 17.100.000 Ergebnisse (0,18 Sekunden)" and "Erweiterte Suche". On the left side, there are filters for "Alles", "Bilder", and "Mehr". The main search results area shows a list of links, with the top result being "Fotos Fiat Punto, Bilder Punto, Fotos Fiat" from a website named "Foto-Galerie Fiat Punto". Below this, there is a link for "Fiat Punto Evo Fotos und Wallpaper".

Erfahrene Sucher wissen, was hier fehlt: Anführungszeichen. Diese kommen an anderer Stelle in diesem Buch zum Einsatz.



Die wichtige Änderung bestand hier darin, das **s** in **Fotos** zu ergänzen. Der Unterschied zwischen einem guten und einem schlechten Ergebnis ist in diesem Fall der Unterschied zwischen Plural und Singular. Sind Sie auf der Suche nach einer Zeichnung von einer Boeing 737? Dann scheint „*construction design*“ Boeing eine gute Wahl. Aber Achtung, Profis sprechen ihre eigene Sprache und leben mitunter in ihrer eigenen Welt. So sehr, dass auf der Zeichnung

oder dem zugehörigen Text nicht die Bezeichnung „*construction design*“ erscheint. Ein Zeichner weiß selbst, dass es eine Konstruktionszeichnung ist, und erwähnt das im Dokument selbst natürlich nicht. Daher können Sie die Arbeit eines Experten nicht einfach mit Hilfe von geläufigen Ausdrücken finden.

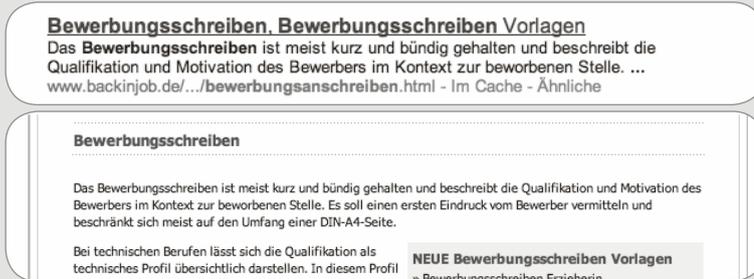
The screenshot shows a Google search interface for the query "construction design" Boeing. The search bar contains the text "construction design" Boeing. Below the search bar, it indicates "Ungefähr 12.900 Ergebnisse (0,23 Sekunden)" and "Erweiterte Suche". On the left side, there are filters for "Alles" and "Mehr". The main search results area shows a list of links, with the top result being "McGraw-Hill Construction Design-Build" from a website named "designbuild.construction.com". Below this, there is a link for "Airports - Northwest Mountain Region - Engineering, Design, and Construction - Design Resources" from the FAA website.

Mit „construction design“ Boeing landen Sie bei einer Passage aus diesem Buch und nicht bei einer Konstruktionszeichnung einer Boeing.

Noch ein Beispiel. Sie suchen in Google echte Bewerbungsschreiben, keine Muster. Welche Frage stellen Sie dann? Meistens bekomme ich diese Antwort:



Sie landen bei einer Site mit Musterbewerbungen, nicht mit echten:



Vielleicht sagen Sie: „Ich suche einfach eine Site mit Bewerbungen, dann ist das doch prima!“ Natürlich, aber Ihr Ziel war es, ein echtes Bewerbungsschreiben zu finden. Denken Sie erneut nach. Was funktioniert besser? Warum ist „Bewerbungsschreiben“ nicht klug, wenn Sie nach Bewerbungsschreiben suchen?

Haben Sie je Folgendes in einer Bewerbung geschrieben:

„Anbei sende ich Ihnen eines meiner Bewerbungsschreiben.“?

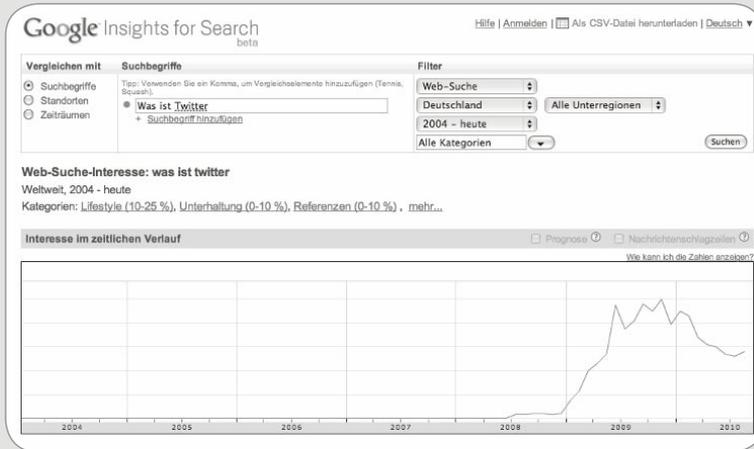
Warum nicht? Wer ein Bewerbungsschreiben versendet, sagt im Bewerbungsschreiben nicht, dass es ein Bewerbungsschreiben ist. Das wissen die Empfänger bereits. Kurzum, ein Bewerbungsschreiben *existiert* einfach. In dem Brief selbst wird darauf nicht aufmerksam gemacht. In einem Gedicht steht meistens nicht das Wort *Gedicht*. Auf einem Etikett steht nicht, dass es ein *Etikett* ist. Gedichte und Etiketten sind uns geläufig.

Und doch erwarten Sie, dass der Suchbegriff „Bewerbungsschreiben“ in einem Bewerbungsschreiben steht. Die Suchmaschine hat Sie enttäuscht. Sie meckern über das Ding und sagen: „Was ich suche, gibt es einfach nicht, ich gebe also auf.“ Sie haben jetzt keine blasse Ahnung mehr, wo Sie suchen müssen.

Kein Wunder. Sie suchen nach Begriffen, die sich nicht im gesuchten Dokument befinden. Keine Suchmaschine der Welt kann Ihnen dabei helfen. Wie Sie mithilfe des Google-Codes trotzdem ein echtes Bewerbungsschreiben finden, lesen Sie auf Seite 119.

Benutzen Sie den „Google-Code“. Gehen Sie nicht von dem aus, was Sie suchen, sondern von dem, was Sie finden möchten. Stellen Sie keine Frage, sondern geben Sie die Antwort. Stellen Sie sich vor, was Sie suchen. Versuchen Sie logisch zu kombinieren, wie die gesuchte Information formuliert sein könnte. Welche Begriffe verwendet jemand, um meine Frage zu beantworten? Der beste Weg, um die richtigen Worte zu finden, ist so zu tun, als ob Sie selbst der Verfasser der Antwort wären.

Ein Beispiel: Ich nutze den Google-Code für die Frage: „Was ist Twitter?“



Dieses Bild zeigt, das die breite Masse erst seit 2009 kräftig danach gesucht hat, was es mit Twitter auf sich hat.

Laut Google-Statistik geben die meisten Benutzer einfach die Frage *Was ist Twitter* ein. Ist das die beste Lösung? Sie fragen Google nach allen Quellen, die die Wörter *was ist Twitter* auf einer Seite haben. Aber was ist, wenn der Autor der Website andere Begriffe verwendet? Wenn er z.B. schreibt, wie Twitter *funktioniert*?

Auch damit kommen Sie nicht weiter, sondern landen auf einer Seite, auf der sich andere fragen, wie Twitter funktioniert.

The screenshot shows a Google search result for the query 'funktioniert Twitter'. The search bar contains the text 'funktioniert Twitter' and a 'Suche' button. Below the search bar, it indicates 'Ungefähr 3.120.000 Ergebnisse (0,13 Sekunden)' and 'Erweiterte Suche'. The search results include a snippet from a page titled 'wie funktioniert TWITTER?' dated 25. Apr. 2009, with the text: '25. Apr. 2009 ... ich blick bei TWITTER nicht durch.bitte helft mir. ... Hallo, ich stand auch vor kurzem vor dem Problem **Twitter**, man kann ja denke ich nicht länger ... www.gutefrage.net/frage/wie-funktioniert-twitter - Im Cache - Ähnliche'. Below the snippet, there is a link to 'Twitter funktioniert nicht!?? (probleme) - 25. Juli 2010'. On the left side, there are navigation options: 'Alles', 'Videos', 'Diskussionen', and 'Mehr'.

Wagen wir einen zweiten Versuch:

The screenshot shows a Google search interface. The search bar contains the text "Funktionsweise von Twitter". Below the search bar, it indicates "Ungefähr 101.000 Ergebnisse (0,17 Sekunden)" and "Erweiterte Suche". On the left side, there are navigation options: "Alles" (selected), "Mehr", and "Das Web durchsuchen". The search results display a link titled "Twitter Befehle | Alltag Medienzukunft" with a snippet: "Ich möchte gar nicht so sehr auf die Funktionsweise von Twitter eingehen. Diese ist vielerorts nachzulesen. Ich möchte euch vielmehr mitteilen was mich ... www.2-blog.net/2008/twitter-befehle/ - Im Cache - Ähnliche".

Dieses Dokument erläutert eben gerade nicht die Funktionsweise von Twitter.

Zeit für den Google-Code!



Wie würden Sie selbst erklären, wie Twitter funktioniert? Sie müssen nicht die ganze Antwort vorwegnehmen, sondern nur den Anfang. Wahrscheinlich fängt Ihre Antwort so an:
„Mit Twitter können Sie ...“

Den Rest der Antwort kennen Sie zwar nicht, aber genau das ist der Trick beim Google-Code. Sie brauchen nur den Anfang. Sehen Sie sich die folgende Anfrage an:

The screenshot shows a Google search interface. The search bar contains the text "Mit Twitter können Sie". Below the search bar, it indicates "Ungefähr 7.190 Ergebnisse (0,20 Sekunden)" and "Erweiterte Suche". On the left side, there are navigation options: "Alles" (selected), "Mehr", and "Das Web durchsuchen" with the subtext "Seiten auf Deutsch". The search results display a link titled "Zwitscher, triller, flöt – Wie Sie Twitter für sich nutzen können ..." with a snippet: "2. Sept. 2009 ... Mit Twitter können Sie ebenso Markt- oder Sozialforschung betreiben und so mögliche Trends aufspüren. Mit Tweetstats etwa können Sie sich ... karrierebibel.de/zwitscher-triller-flot—wie-sie-twitter-fur-sich-nutzen-koennen/ - Im Cache - Ähnliche".

Stellen Sie keine Fragen, sondern verfassen Sie Antworten.

Warum Anführungszeichen? Damit legen Sie und nicht Google die Reihenfolge der Wörter fest. Ohne Anführungszeichen können die Wörter irgendwo auf der Seite stehen, mit müssen sie exakt in der Reihenfolge vorkommen, in der Sie sie eingeben.

Sie überlegen sich im Voraus einen Teil der Antwort. Wenn Sie das gut machen, bekommen Sie den zweiten Teil geliefert. Ihr Ergebnis ist nicht nur genau, vielmehr nimmt die Anzahl der Ergebnisse rapide ab, wenn Sie Anführungszeichen verwenden und mit mehreren Wörtern suchen.



Der ultimative Google-Code für die Frage „Was ist Twitter“ ist verblüffend einfach:

Google

Ungefähr 58.900 Ergebnisse (0,14 Sekunden) Erweiterte Suche

Alles
 Videos
 Mehr

Twitter – Wikipedia
Twitter ist eine Anwendung zum Mikroblogging. Es wird auch als soziales Netzwerk oder ein meist öffentlich einsehbares Tagebuch im Internet definiert. ...
 Funktionen - Unternehmen - Technologie - Nutzung und Verbreitung
de.wikipedia.org/wiki/Twitter - Im Cache - Ähnliche

Neben *Twitter* verwende ich hier nur noch das Wort *ist*. Das kleine Wort *ist* ist, wenn Sie denken wie ein Autor, eine prima Methode, um Meinungen oder Aussagen über etwas zu finden. Betrachten Sie die folgenden Beispiele:

Google

Ungefähr 25.600 Ergebnisse (0,10 Sekunden) Erweiterte Suche

Alles
 Mehr

Das Web durchsuchen

SD Memory Card – Wikipedia
 Eine SD Memory Card (Kurzform für Secure Digital Memory Card; deutsch Sichere digitale **Speicherkarte**) ist ein digitales Speichermedium, das nach dem Prinzip ...
de.wikipedia.org/wiki/SD_Memory_Card - Im Cache - Ähnliche

Google

Ungefähr 14.700 Ergebnisse (0,15 Sekunden) Erweiterte Suche

Alles
 Mehr

Das Web durchsuchen

Rollator – Wikipedia
 Der **Rollator** ist ein anerkanntes Hilfsmittel der Gesetzlichen Krankenversicherung (GKV). Im Hilfsmittelverzeichnis der GKV ist der Rollator in der ...
de.wikipedia.org/wiki/Rollator - Im Cache - Ähnliche

Google

Ungefähr 202.000 Ergebnisse (0,20 Sekunden) Erweiterte Suche

Alles
 Mehr

Das Web durchsuchen

Après-Ski – Wikipedia
 Die Verbindung von Skisport und **Après-Ski** ist vergleichbar mit der Verbindung von ... Eins der bekanntesten Lieder zum **Après-Ski** ist Schiffoan (1976; dt. ...
de.wikipedia.org/wiki/Après-Ski - Im Cache - Ähnliche

Wie suchen Sie? Studien zeigen, dass die meisten Menschen nicht mehr als drei Wörter eingeben, meistens sogar nur zwei bzw. eins. Hier sehen Sie eine Liste mit den Suchanfragen in Google.de aus den letzten sieben Tagen.

Web-Suche Interesse
Deutschland, Letzte 7 Tage

Suchbegriffe	
Top-Suchanfragen	Zunehmende Suchanfragen
1. facebook	1. mary shelley Aufsteiger
2. youtube	2. sarrazin +300 %
3. ebay	3. schalke 04 +70 %
4. google	4. transfermarkt +70 %
5. wetter	5. kino.to +60 %
6. you	6. lottozahlen +60 %
7. gmx	7. bundesliga +50 %
8. hotmail	8. lotto +50 %
9. amazon	9. deichmann +40 %
10. web.de	10. facebook.login +40 %

[Diese Tabelle einbetten](#) [Diese Tabelle einbetten](#)

Mitunter schleichen sich Tippfehler in die Suche ein, doch Google löst das Problem derzeit unmittelbar bei der Eingabe. Sie erhalten den Vorschlag, die Schreibweise des Worts zu ändern, und sehen außerdem schon die ersten zwei Websites, auf denen der Begriff in der richtigen Schreibweise vorkommt.

Google parkinson

Ungefähr 1.020.000 Ergebnisse (0,27 Sekunden) Erweiterte Suche

Meinten Sie: [parkinson](#) Die ersten 2 angezeigten Ergebnisse

[Parkinson-Krankheit – Wikipedia](#)
Die **Parkinson-Krankheit** bzw. Morbus **Parkinson** (weitere Synonyme: Idiopathisches **Parkinson-Syndrom** (IPS), Parkinsonsche Krankheit, ältere Bezeichnung: ...
Einklassung - Geschichte - Erkrankungsalter und Häufigkeit
de.wikipedia.org/wiki/Parkinson-Krankheit - Im Cache - Ähnliche

[Parkinson \(Morbus Parkinson, Parkinson-Krankheit\) - Onmeda ...](#)
Morbus **Parkinson** zählt zu den häufigsten Krankheiten in der Neurologie und ...
Charakteristisch für Morbus **Parkinson** ist ein fortschreitender Verlust von ...
www.onmeda.de > Krankheiten - Im Cache - Ähnliche

Ergebnisse für: **parkinson**

Alles
 Videos
 Mehr

Das Web
Seiten auf Deutsch
Seiten aus Deutschland

Alle
Letzte 2 Tage
 Mehr Optionen

Wer schnell klickt, sieht nicht einmal, dass seine Suchanfrage fehlerhaft war.

Hier möchte jemand mehr über die Parkinson-Krankheit wissen. Meistens hat ein Suchender jedoch eine weitere Absicht. Angenommen, die tatsächliche Frage lautet: Wie tödlich ist Parkinson? Was wäre dann eine gute Suche?

Google Parkinson lebensbedrohlich

Ungefähr 79.800 Ergebnisse (0,45 Sekunden) Erweiterte Suche

[Parkinson HEXAL | Alles über Morbus Parkinson, Symptome und Behandlung](#)
Morbus **Parkinson** ist nicht **lebensbedrohlich**, und erfreulicherweise sind leichtere Verlaufsförmungen häufiger als schwere. Je früher die Diagnose gestellt wird ...
www.parkinson.hexal.de/ - Im Cache - Ähnliche

Alles
 Mehr

Das Web
Seiten auf Deutsch

Das oberste Ergebnis ist nicht relevant.



Wenden Sie jetzt den Google-Code an. Wie würden Sie eine Antwort beginnen? Es sind nur zwei Antworten möglich: Die Krankheit ist tödlich oder die Krankheit ist nicht tödlich. Geben Sie das in Google ein:

The screenshot shows a Google search interface. The search bar contains the text "Parkinson ist tödlich". To the right of the search bar is a "Suche" button. Below the search bar, the text "Erweiterte Suche" is visible. On the left side, there are links for "Alles" and "Diskussionen". The main search results area displays a message: "Keine Ergebnisse für 'Parkinson ist tödlich' gefunden."

Die Suche liefert kein Ergebnis. Ist die Krankheit etwa nicht tödlich? Fragen Sie Google:

The screenshot shows a Google search interface. The search bar contains the text "Parkinson 'nicht tödlich'". To the right of the search bar is a "Suche" button. Below the search bar, the text "Ungefähr 350 Ergebnisse (0,13 Sekunden)" and "Erweiterte Suche" are visible. On the left side, there are links for "Alles" and "Mehr". Under "Das Web", there are links for "Seiten auf Deutsch", "Seiten aus Deutschland", and "Mehr Optionen". The main search results area displays two results:

- Parkinson: Nebensymptome**
Normalerweise ist die Parkinson-Erkrankung **nicht tödlich**. Die Betroffenen haben bei guter medikamentöser Behandlung meistens eine nahezu gleich hohe ...
[parkinson.gesund.org/.../nebensymptome.htm](#) - Im Cache - Ähnliche
- Morbus Parkinson: Therapie und Verlauf - Chronische Krankheiten ...**
Morbus Parkinson führt jedoch in den meisten Fällen zur Pflegebedürftigkeit. Die Krankheit selbst ist **nicht tödlich**, allerdings treten im Laufe der ...
[www.blog.de/tb/al/r/...parkinson.../6470644/](#) - Im Cache - Ähnliche

Ja, so ist es. Die offizielle Parkinson-Vereinigung erklärt genau das auf ihrer eigenen Site. Alles, was Sie getan haben, um die Antwort zu finden, ist, die Antwort selbst zu geben.

Angenommen, Sie suchen eine *Zugverbindung in die Bretagne*. Im letzten Kapitel lernen Sie noch ein besonderes Werkzeug kennen, mit dem Sie den Google-Code noch präziser auf Ihre Frage abstimmen. Ich zeige Ihnen vorab schon, wie Sie nur in Überschriften im Titel der Webseite suchen und damit nicht im Text selbst. Das machen Sie mit **allintitle**:

The screenshot shows a Google search interface. The search bar contains the text "allintitle: Bahn Bretagne". To the right of the search bar is a "Suche" button. Below the search bar, the text "Ungefähr 23 Ergebnisse (0,18 Sekunden)" and "Erweiterte Suche" are visible.

Der Erfolg hängt von der richtigen Wortwahl ab: Mit **Zug** statt **Bahn** geht es von der Schweiz aus in die Bretagne.



Da es in der Überschrift steht, wird der Autor wahrscheinlich im restlichen Text darüber schreiben. Sie erwarten irgendwo im Web eine Seite, in deren Titel eine Überschrift mit den Wörtern **Bahn** und **Bretagne** steht. Und das stimmt:



Bahnreise Bretagne mit TGV - komfortabel und schnell

Bretagne

Erholbarer Natururlaub in den schönsten Landstrichen der Bretagne



Google-Anzeigen

Bahn-Anreise Bretagne

Bahn-Anreise Bretagne per TGV Thalys: Für die Bahn-Anreise Bretagne per TGV- und Thalys-Zügen finden Urlauber wichtige Reise-Informationen.

Reiseinformationen zum Bahn fahren in Frankreich und Bretagne bei Bretagne-Netz sind u.a. TGV Strecken und TGV Bahnhöfe Bretagne, Thalys Strecken und Verbindungen, Bahntickets, Fahrkarten TGV online, Bahnfahren mit Handicap und Bahnfahren mit Gepäck (Fahrräder, Kinderwagen, Surfbrett)

TGV Bahnstrecken und TGV Bahnhöfe Bretagne

Der TGV oder auch Train Grande Vitesse ist nicht nur der schnellste Zug auf den Bahnstrecken in der Bretagne, Frankreich bis nach Deutschland. Der schnelle und supermoderne Zug hat im Gegensatz zu den ICE-Zügen bisher auch keine schwerwiegenden technischen Probleme, im Gegenteil.

Das Streckennetz für TGV-Zugverbindungen wird ständig erweitert und modernisiert und die Angebote für Bahnfahrten in die Urlaubsziele der Bretagne werden dank des TGV ständig erweitert.

Etwas exotischer ist die Suche nach einer Strecke in Marokko:

Google

Ungefähr 119.000 Ergebnisse (0,23 Sekunden) Erweiterte Suche



Keine schlechte Suche, aber mit dem Google-Code könnte sie besser sein. Welcher Passus kommt in nahezu allen Wegbeschreibungen vor? Keine Idee? Macht nichts. Öffnen Sie ein paar andere Wegbeschreibungen in Google und suchen Sie etwas, das in allen Dokumenten wiederkehrt. Es ist nicht immer das Wort **Strecke**.

Was jedoch fast überall erwähnt wird, ist „von ... nach ...“ z.B. von Marrakesch nach Essaouira. Geben Sie das also ein. Benutzen Sie Englisch, um die Anzahl der Treffer zu erhöhen:

The screenshot shows a Google search interface. The search bar contains the text "Marrakech to Essaouira" and a "Suche" button. Below the search bar, it indicates "Ungefähr 75.700 Ergebnisse (0,31 Sekunden)" and "Erweiterte Suche". On the left side, there are navigation options: "Alles" (selected), "Maps", and "Mehr". Below that, there is a section "Das Web durchsuchen" with "Seiten auf Deutsch" and "Mehr Optionen". The main search results are as follows:

- Essaouira, Morocco Travel Forum - VirtualTourist.com** - [Diese Seite übersetzen]
30 Beiträge - 26 Autoren - Letzter Eintrag: Vor 5 Tagen
Travelling late in the day from **Marrakech to Essaouira**. Posted: Wed September 3, 2008 01:31 PM UTC. Our flight arrives at 5pm on oct ...
forum.virtualtourist.com/forum-2113916-1-Travel-Essaouira-1-forum.html - Im Cache - Ähnliche
- Essaouira - Wikipedia, the free encyclopedia** - [Diese Seite übersetzen]
The road from **Marrakech to Essaouira** is a straight line, explaining the King's choice of this port among the many that the Moroccan coast offers. ...
en.wikipedia.org/wiki/Essaouira - Im Cache - Ähnliche
- Map of the medina of Essaouira, transport Marrakech - Essaouira**
... - [Diese Seite übersetzen]
Airlines and for transport from **Marrakech to Essaouira** : a bus compagny Supratour. Map of Essaouira medina.
www.ryad-watier-maroc.com/riad-essaouira-informations-34.html - Im Cache - Ähnliche
- Bus from Marrakech to Essaouira - Marrakech Message Board**
... - [Diese Seite übersetzen]
9 Apr 2010 ... On TripAdvisor's Marrakech travel forum, travellers are asking questions and offering advice on topics like "Bus from **Marrakech to** ...
www.tripadvisor.co.uk/ShowTopic-g293734-t9196-k3536250-Bus_from_

Jetzt kennen Sie den Google-Code. Im letzten Kapitel wird dieses Thema vertieft. Aber als Nächstes lernen Sie einige Mythen über Google kennen und erfahren, was es außer Google noch gibt.

2.

Google- Geheimnisse

In diesem Kapitel räume ich mit einigen **hartnäckigen Irrtümern über Google** auf und enthülle, **wie Google zu den Ergebnissen kommt**. Lernen Sie das blühende Leben jenseits von Google kennen und entdecken Sie, in welcher baden-württembergischen Stadt am häufigsten nach dem Ausdruck Kokain gesucht wird.

$$R'(u) = c \sum_{v \in B_u} \frac{R'(v)}{N_v} + cE(u)$$

Das Geheimnis von Google. Mit dieser Zauberformel verdienen die Eigentümer der amerikanischen Suchmaschine Milliarden von Euro pro Jahr.

Welche der folgenden Behauptungen sind wahr?

1. Google ist gut im Finden neuer Sites.
2. Google ist vollständiger als andere Suchmaschinen.
3. Google ist derzeit die beste Suchmaschine.
4. Suchergebnisse von Google sind immer zuverlässig.
5. Google sucht weltweit.

Die Antwort ist: Alle fünf Aussagen sind **nicht wahr**.

Wie kann das sein? Um das zu verstehen, müssen Sie mehr darüber wissen, wie Google funktioniert. Das ist ein kompliziertes Thema. Daher beginnen wir am Anfang: Wir tun einfach so, als ob das Internet heute neu erfunden wird.



Um die Arbeitsweise von Google zu verstehen, tun wir so, als ob das Internet ganz neu wäre. Was wären Ihre ersten Wörter im Web?

Sie haben die Ehre, die erste Website der Welt zu füllen. Worüber werden Sie schreiben? Wird es eine Site über die politischen Verhältnisse zwischen Ost und

West? Eine Webzeitung mit aktuellen Internetnews? Ein Weblog mit Fotos vom ersten Geburtstag Ihres Kindes? Oder eine Site über „Lesen“? Ja, das wird es: eine Site über das Lesen. Sie eröffnen heute eine Site über *Bücher*. Sie sagen, welche Bücher Sie gut finden und warum. Einen Tag später kommt Ihr Nachbar auf die gleiche brillante Idee. Im neuen Web gibt es jetzt zwei Sites zum Thema Lesen, mehr nicht. Am dritten Tag wird *Startseite.de* gegründet. Unter dem Titel „Schöne Lesesites“ platziert *Startseite* einen Link auf Ihre Site und nicht auf die Ihres Nachbarn. *Startseite* führt allerdings eine teure Werbekampagne mit ihm durch, die sofort zehntausende Besucher anzieht. Sie haben viel weniger Besucher. Am Tag vier kommen allerlei andere Sites ins Web, aber keine neue Site über das Lesen. Am Tag fünf wird Google eingerichtet. Die amerikanische Suchmaschine versucht, den Inhalt des Webs auf einer Festplatte zu speichern. Das ist eine einfache Aufgabe, da nur wenige Dutzend Websites vorhanden sind: *Startseite*, Ihre Site, die Ihres Nachbarn und noch ein paar andere Sites (die nicht vom Lesen handeln).



Der Keller von Google in der Stanford-Universität 1996 (links). Google hieß damals noch Backrub. Die Farben des Google-Logos stammen von den Legosteinen (rechts).

Wie funktioniert Google?

Was passiert nun, wenn jemand *Lesen* in Google eingibt? Welche Site steht dann an erster Stelle? Ist es Ihre, weil Sie einen Link von *Startseite* bekommen haben? Ist es *Startseite*, weil diese Site am häufigsten besucht wird? Oder ist es die Website Ihres Nachbarn? Hier sehen Sie die Antwort:

The screenshot shows a Google search interface. The search bar contains the word "Lesen". Below the search bar, it indicates "3 Ergebnisse (0,06 Sekunden)" and "Erweiterte Suche". On the left side, there are navigation options: "Alles" (selected), "Bilder", "Bücher", and "Mehr". Below these are filters: "Das Web durchsuchen" (Seiten auf Deutsch), "Alle" (Letzte 3 Tage), and "Standardansicht". The search results are as follows:

- 1. Result:** [Meine Website über das Lesen](#)
Ich halte sehr viel vom **Lesen**. Ein gutes Buch stimuliert die Lesefreude. Sowohl auf Deutsch als auch in anderen Sprachen. Ein gutes Buch ist wie...
www.diebestesiteüberslesen.de/ - Im Cache - Ähnliche
- 2. Result:** [Die Site meines Nachbarn über das Lesen](#)
Lesen finde ich anstrengend... **Lesen** ist eher etwas für Leute, die ...
www.nachbarkannauchlesen.de/?age_group_id=0...item... - Im Cache - Ähnliche
- 3. Result:** [Bücher.Startseite.de - Alles über Bücher, Autoren und Bestseller](#)
Stiftung **Lesen** - Verlegerverband - Fachausbildung in der Buchbranche - Verband freier Verleger - Beschluss fester Buchpreise ...
bücher.startseite.de/ - Im Cache - Ähnliche

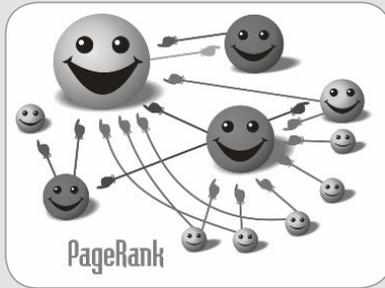
Sie sehen, dass Sie auf Platz 1 stehen, Ihr Nachbar auf Platz 2 und die viel bekanntere Startseite auf Platz 3. Wie kommt das? Betrachten wir die Logik, die hinter einer Ergebnisliste in Google steht.

Die wichtigste Site ist für Google vor allem diejenige, über die die Leute reden. Das passiert im Internet über Links. Wenn eine Website einen Link auf eine andere Website platziert, sieht Google das als eine Empfehlung an („geh da mal gucken“). Google findet es gut, Sites mit vielen Links in der Ergebnisliste höher zu platzieren als Sites, über die niemand spricht. In unserem Beispiel landen Sie daher auf Platz 1, denn Sie haben die einzige Site mit einem Link. Niemand spricht über *Startseite* und niemand verlinkt zu Ihrem Nachbar. Nur *Startseite* hat Sie verlinkt. Das sieht Google als Empfehlung, Sie auf Platz 1 zu setzen. *Startseite* selbst landet auf Rang zwei, da Google auch berücksichtigt, wie häufig sich das gesuchte Wort auf einer Site befindet. Steht das Wort *Lesen* nur ein paar Mal auf einer Site, bedeutet das eine niedrigere Relevanz als bei einer Site, die das Wort häufiger enthält (Das Wort *Lesen* tausendfach auf einer Site unterzubringen ist übertrieben, denn Google betrachtet das als Spam.). Ihr Nachbar befasst sich auf seiner Site nur mit Lesen. *Startseite* behandelt hingegen noch allerlei andere Themen.

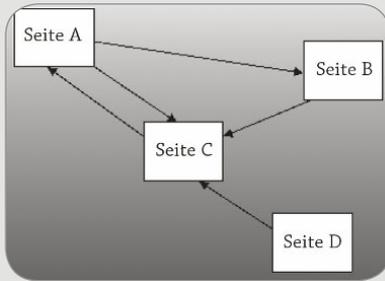
Google hat also eine eigene Methode, Suchergebnisse zu gewichten. In der Regel besetzen nicht die besten oder neuesten Websites die Spitze der Ergebnisliste, sondern die am häufigsten beachteten Sites. Google glaubt an das Funktionieren des sozialen Netzwerks.

Die Idee des früheren niederländischen Gewerkschaftsvorsitzenden der NKV (später mit der NVV in die FNV aufgegangen) war, dass die Wirtschaft der Niederlande durch zweihundert Menschen geführt wird, die einander kennen

und beeinflussen. Google arbeitet nach einem ähnlichen Prinzip. Die Idee, danach zu schauen, wie viele Links auf eine Site verweisen, hat einen Namen: *PageRank*.

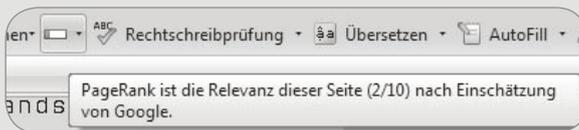


Google schaut nicht nur, ob man über Sie spricht, sondern auch, wer das tut. Links von wichtigen Sites fallen mehr ins Gewicht. Es hat keinen Sinn, mit Freunden abzusprechen, sich untereinander mehrfach zu verlinken, um in der Ergebnisliste von Google höher zu steigen (es sei denn, dass man bedeutende Freunde hat).



Vier Websites. Welche spricht Google am meisten an? Es ist Seite C, da sie die meisten Links erhält.

Die Suchmaschine betrachtet, wie häufig eine Website von anderen erwähnt wird und merkt sich das sehr genau. Je häufiger auf eine Website verwiesen wird, desto wichtiger wird diese. Der PageRank ist eine Note von 0 bis 10. Die wichtigsten Sites haben einen PageRank von 10. Neulinge bekommen eine 0.



Laden Sie von toolbar.google.com die Google Toolbar herunter und aktivieren Sie unter Optionen/Tools den Eintrag PageRank. Ab jetzt sehen Sie, wenn Sie sehr genau schauen, einen grünen Balken bei einem PageRank von 1 oder mehr und einen weißen Balken bei 0. Bewegen Sie den Mauszeiger ohne zu klicken über den Balken, wird die genaue Zahl angezeigt. Je höher die Zahl, desto mehr Menschen verlinken zu der Site.

Plagegeister missbrauchen das. Sie platzieren beispielsweise die Wörter „komische Frisur“ auf verschiedenen Websites, gefolgt von einem Hyperlink auf den Lebenslauf von Volker Bouffier. Google denkt: Jemand sucht nach „komische Frisur“, welche Website wird in diesem Zusammenhang am häufigsten genannt? Antwort: die offizielle Regierungssite. Und dass, obwohl die Wörter „komische“ und „Frisur“ nicht ein einziges Mal auf der Regierungssite vorkommen! Monatelang führte diese Art von „Google Bombe“ zur Erheiterung. Inzwischen hat Google die Suchergebnisse geändert.



Wer ist ein großer Versager? Aufgrund der zahlreichen Links, die er im Zusammenhang mit den Begriffen „miserable failure“ erhielt, war dies jahrelang George Bush.

Scientology versteht es, kritische Veröffentlichungen über die Bewegung aus den Top 3 von Google zu verdrängen. Die Sekte war nicht erfreut über die Warnungen der niederländischen Autorin Karin Spaink vor Scientology. Daher griff sie zu einem Trick und errichtete weltweit über zweihundert „Linkfarmen“, auf denen das Wort *scientology* immer auf positive Texte verwies. Die Kritik von Spaink sackte immer weiter weg.

Wer sind die „200 von Google“ in Deutschland? Google platzierte beispielsweise Mister Wong in der Hierarchie oft als Nr. 1. Danach folgen die deutsche Google-Site, Universitätssites sowie mit Spiegel, Zeit und FAZ wichtige Mediensites. Je älter die Site, desto höher ist meist der PageRank. Warum? Ein Baby hat erst wenig Freunde, ein Erwachsener hat jedoch allerlei Bekannte. Je länger Sie leben, desto größer ist die Chance, dass andere Sie kennen. Gutes Abschneiden in Google ist tatsächlich eine Frage von gutem Sitzfleisch.

Voelspriet entschlüsselt das PageRanking von Google

Marco Derksen (Upstream) Sonntag, 12. Januar 2003, 21:11

Rubrik: Suchmaschinenmarketing | Gelesen: 1241 mal (2568 mal angeklickt)

Die Suchseite Voelspriet hat das PageRanking von Google entziffert. Praktische Anwendung dieses Funds ist die "Google Top 100 Niederlande" mit den wichtigsten niederländischen Meinungsführern.

Der Herausgeber Reed Elsevier ist laut Google mit "Elsevier Science" die wichtigste Website der Niederlande. Es ist die einzige Website in den Niederlanden, die 9 von 19 PageRanking-Punkte hat. In Belgien ist die Europäische Kommission für Standardisierung die Nummer 1 (7 von 10).

Voelspriet, selbst auf Platz 491, arbeitete Wochen an der Entzifferung des PageRankings, dessen mathematische Formel übrigens kein Geheimnis ist. Die Top 100 werden ab dem 1. Februar jeden Monat auf Voelspriet veröffentlicht, die Top 1000 sind nur auf Anfrage erhältlich. Die Top 100 wird auch für Belgien erhältlich sein.

Ich versuche schon seit vielen Jahren, die Systematik von Google zu durchschauen. Inzwischen ist Google selbst die wichtigste Website für Google.

Google kennt insgesamt zweihundert Kriterien, um den Platz in der Ergebnisliste festzulegen. Eine Site, die häufig aktualisiert und von anderen immer wieder neu verlinkt wird, gewinnt gegenüber einer enzyklopädischen Site, die im Web nicht oft erwähnt wird. Es wird auch nachgesehen, wie oft das gesuchte Wort im Domainnamen oder im Link steht. In Deutschland werden jährlich etwa 1,2 Milliarden Euro dafür ausgegeben, Websites so zu optimieren, dass Google sie maximal bewertet.

Sie wissen nun, wie Google funktioniert. Zurück zu den fünf Behauptungen. Warum sind sie falsch?

Google sieht neue Sites nicht schnell genug

Es dauert eine Weile, bis Menschen auf neue Sites verweisen. Dennoch finden Sie in Google Verweise auf Nachrichtenmeldungen. Suchen Sie in www.google.de nach *Tag der deutschen Einheit*:



Ungefähr 691.000 Ergebnisse (0,09 Sekunden)
Erweiterte Suche

Das vierte Ergebnis sieht aktuell aus: „vor 13 Stunden gefunden“.

News zu Tag der Deutschen Einheit

"Es ist gewaltig, was sich getan hat!" - vor 13 Stunden gefunden
 So ist es nicht verwunderlich, dass der SPD-Ortsvereinsvorsitzende den **Tag der Deutschen Einheit** als zentral betrachtet: "Dieser Tag wird vielen Deutschen ...
 Südwest Presse
 CDU will Volksfest am **Tag der Deutschen Einheit** - WELT ONLINE - 3 weitere Artikel »
 Zum Feiern verreisen: Von der **deutschen Einheit** bis Jim Knopf - sueddeutsche.de - 9 weitere Artikel »

Die „News zu Tag der Deutschen Einheit“ stammen von den Nachrichtensuchern von Google, Google news Deutschland (news.google.de). Ihre Suchfrage wird jeweils mit den Artikeln von Hunderten von Nachrichtenmedien verglichen. Aber Vorsicht: *News finden* ist etwas anderes als *neue Sites finden*.

Proteste gegen Softwarepatente in Europa

Nach dem Vorbild der USA stimmt die Europäische Kommission heute über die Nichteinführung von Softwarepatenten in der Europäischen Union ab. Diese beinhalten, dass nicht nur ganze Programme, sondern auch Ideen patentiert werden können.

Referenzen und weitere Informationen

Die offizielle Protestsite

Eine Ankündigung einer neuen Website wurde nicht ins Web gesetzt. Den Link auf die Protestsite sieht Google nicht direkt.

Das Weblog Sargasso verweist auf *Die offizielle Protestsite*. Das ist eupat.ffii.org. Dort können Sie gegen die Einführung von Softwarepatenten in Europa protestieren. Die Site ist jetzt nicht mehr neu, war es aber, als Sargasso darüber schrieb. Was hat Google in der Woche gesehen, in der die Protestsite gestartet wurde? Überhaupt nichts. Google sieht die Neulinge nicht:

The screenshot shows a Google search interface. The search bar contains the text "site:http://eupat.ffii.org/". To the right of the search bar is a "Suche" button. Below the search bar, there is a link for "Erweiterte Suche". On the left side, there are filters for "Alles" and "Mehr". The main results area displays the message: "Es wurden keine mit Ihrer Suchanfrage - site:http://eupat.ffii.org/ - übereinstimmenden Dokumente gefunden."

Mit der Suchanfrage „site: (name einer website)“ sehen Sie, wie viele Seiten Google von einer beliebigen Website gespeichert hat. In diesem Fall: keine.

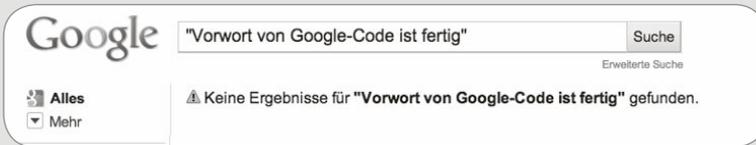
Was ist geschehen? Die Website eupat.fi.org ist noch so neu, dass Google den Inhalt nicht kennt. Es dauert eben, bevor die Suchmaschine die Website sieht.

Und bis dahin tappen Sie im Dunkeln. Eine der Ursachen ist, dass Google mangelhafte Links von Sites findet, die den Neuling empfehlen. Eine neue Site hat es daher zu Beginn schwer, aufzufallen. Manchmal dauert es eine Woche oder länger, bevor eine neue Site in Google zu sehen ist.

Realtime-Web ist damit die Achillesferse von Google, also die Fähigkeit, alles zu sehen, was soeben ins Web gestellt wurde. Normale Suchmaschinen können da nicht mitkommen. Ich schreibe z.B. morgens auf Twitter, dass das Vorwort von diesem Buch fertig ist.



Am selben Abend hat Google meine Nachricht noch nicht bemerkt.



Google besucht weltweit nur die beliebtesten Websites mehrmals am Tag. Derzeit versucht die Suchmaschine soziale Netzwerke häufiger zu besuchen. Seit Januar 2010 werden z.B. Tweets von Twitter täglich mitgenommen. Das geht den Kennern aber noch nicht schnell genug. Google sieht Millionen Websites nicht täglich. Es sieht nicht Ihren Nachbarn, aber Ihren entfernten Freund. Mit blogsearch.google.de probiert die Suchmaschine, wieder an Boden zu gewinnen:



Google Blogsearch ist ein Versuch von Google, dem Web in Echtzeit zu folgen ...

Aber eine Stunde nach einer Rede von Volker Bouffier kann Google noch immer nichts finden:



... aber mit wechselndem Erfolg.

Wie finden Sie dann trotzdem das *Realtime-Web*? Probieren Sie für Twitter die Suchmaschine search.twitter.com aus oder sehen Sie sich www.scoopler.com an:



Scoopler versucht, so viel Live-Information anzuzeigen wie möglich.

Es funktioniert noch nicht perfekt, aber es ist ein Anfang.

Praktisch ist auch IceRocket, das Sie unter www.icerocket.com erreichen. Wenn Sie *Big Buzz* wählen, finden Sie allerlei Sites, die Google manchmal erst nach ein paar Tagen oder noch später anzeigen kann. Mit *Auto refresh* legen Sie fest, ob Sie jede Minute, alle drei oder alle fünf Minuten auf dem Laufenden gehalten werden möchten.

The screenshot shows the Icerocket search engine interface. At the top, there are navigation links for Blogs, Web, Twitter, MySpace, News, Images, and Big Buzz. A search bar contains the text 'henk van ess' and a 'Search' button. The Icerocket logo is in the top right corner. Below the search bar, it says 'Big Buzz' and '1 - 50 results for henk van ess'. The results are categorized into 'Blogs' and 'Twitter'. The 'Blogs' section lists several results from 'Roud Questions' and 'Skyring Roud Questions', all mentioning 'Henk van Ess'. The 'Twitter' section shows tweets from 'eurlings' and 'lobsterte' mentioning 'The henk van ess Daily is out'.

Auch IceRocket zeigt Inhalte von Blogs und sozialen Netzwerken nahezu live an.

Google ist nicht vollständig

Google scheint vollständig zu sein. Meistens erhalten Sie mehr Ergebnisse, als Ihnen lieb ist. Aber wie viele Antworten sind das ... genau?

Auf der Homepage von Google, Yahoo oder MSN (jetzt Bing) stand jahrelang die Anzahl der Dokumente, aus denen geschöpft werden konnte. Suchmaschinen standen in einem Wettlauf um den Titel „Wer ist der Größte“. Der Streit ist beigelegt. Zahlen über die Größe werden nicht mehr oder kaum noch angegeben. Mit einem kleinen Trick ist es trotzdem möglich, sich ein Bild zu machen:

The screenshot shows a Google search result. The search bar contains the text 'inurl:http://'. Below the search bar, it says 'Ungefähr 25.830.000.000 Ergebnisse (0,17 Sekunden)'. To the right of the search bar is a 'Suche' button. Below the search bar, there is a link to 'Erweiterte Suche'.

*Fast jeder Link enthält **http://**. Mit **inurl:** suchen Sie nach dem Inhalt eines Links.*

*Mit **inurl:http://** erhalten Sie ein gutes Bild davon, wie viele Dokumente Google sieht.*

Das sind hier mehr als 25,83 Milliarden. Zwei Tage später waren es 25,82 Milliarden.

Und gestern wieder 25,83. Kurzum, Zahlen dieser Art veralten schnell.

Warnung: Hier folgen entmutigende Zahlen:



Mit knapp 26 Milliarden Quellen (Stand 09/2010, ohne Doppelungen) scheint Google vollständig zu sein. Angenommen, dass Sie jedes Dokument innerhalb von einer Sekunde lesen könnten, benötigten Sie 298.958 Tage (= 819 Jahre), um alles einmal zu sehen. Da Sie vermutlich nicht älter als 100 Jahre werden, können Sie nie mehr als 12,2 % ($365 \times 100 : 298.958$) von Google sehen. In einem Menschenleben können Sie also in Google nur an der Oberfläche kratzen. Zum Lesen eines Dokuments ist eine Sekunde sehr wenig, eine Minute scheint realistischer. In diesem Szenario benötigten Sie 49.140 Jahre und sähen nur 0,2 % des Inhalts von Google. Wahrscheinlich sehen Sie noch weniger, denn Sie beginnen natürlich nicht gleich mit null Jahren zu lesen und hören möglicherweise schon vor Ihrem Hundertsten auf. Dadurch wird die Differenz zwischen dem, was Sie sehen können, und dem, was Google sehen kann, noch größer. Frustrierend?

Ja, und dabei ist das erst der Anfang. Google sieht sehr viel, aber lange nicht alle Informationen. Schlimmer noch, die Suchmaschine zeigt nur einen Ausschnitt des Webs. Nach neuen Schätzungen sieht Google nicht mehr als 25 % des öffentlichen Internets. Andere Studien zeigen, dass es halb so schlimm ist und Google etwa 50 Prozent sieht. Mit Sicherheit finden Sie also eine von zwei Quellen nicht mit Google und wahrscheinlich entziehen sich sogar drei von vier Quellen Ihrer Wahrnehmung.

Index

- allintitle: 19, 127
- Alltheweb 35
- Anführungszeichen 16, 68
- Antwort kontrollieren 76
- Antwort vorhersagen 15
- Archivsuche 88
- Ausgangssprache 107, 109

- Bekannte Quellen 67
- Beste Suchmaschine 35
- Bing 36

- Chillingeffects.org 58
- Cluster-Suchmaschinen 41

- Dateityp 116
- Datum wählen 89
- Dogpile 37
- Dokument vorstellen 116
- Domainfilter 100, 101
- Domainnamen 100
- Drei Pünktchen 74

- Englisch 107
- Entfernte Seiten 55
- Exalead 42

- filetype 117
- Freebase 42
- Funktionsweise von Google 25

- Geotargeting 59
- Google Bilder 112
- Google Bücher 82
- Google Maps 109
- Google News Archiv 88
- Google Squared 44
- Google Translate 107

- Hakia 38

- Icerocket 32
- inurl: 121, 129
- Irrtümer über Google 24
- Ist 17

- Juristische Aktionen 55

- Kontrolltrick 76

- Manipulierende Webmaster 51
- Mashup-Suchmaschinen 47
- Menschliche Suchmaschinen 45
- Metasuchmaschinen 37
- Minuszeichen 84

- Neue Sites 29

- Open Directory 46
- Ordnerstruktur 121

- PageRanking 27
- Plural 13, 108

- Quellen, bekannte 67

- Reihenfolge der Wörter 16

- search.twitter.com 32
- Seite wurde entfernt 55
- Semantische Suchmaschinen 38
- Singular 13
- site: 92
- Spezialsuchmaschinen 38, 51
- stopbadware.org 53
- Suchen in Titeln 127
- Suchmaschine
 - Cluster 41
 - mashup 47
 - menschlich 45
 - semantisch 38
 - vergleichen 35

- Technorati 51
- Titel 127
- TLD 100
- Twitter 46

- Überschrift 20

- Vollständigkeit von Google 33

- Webmaster 51
- Weltweite Suche 58
- Wikia 46
- Wolfram Alpha 43
- Wörter ausschließen 84
- www.scooper.com 32

- Yahoo 36
- Yippy 42

- Zensur 55
- Zuverlässigkeit
 - Google 51
 - verbessern 98
 - von Quellen 66, 76