

## Wie entstehen eigentlich die Google-Listen?

# Das ungelüftete Geheimnis der berühmtesten Suchmaschine der Welt

Wer googelt, der findet. Blitzschnell holt diese Suchmaschine aus den weiten Welten des Internets die Antworten auf alle Fragen. Nicht nur eine, sondern ganze Listen von sogenannten Treffern. Manchmal Hunderte, manchmal aber auch Millionen. Aber immer sortiert und durchnummeriert. Wie macht Google das eigentlich?

Philipp Sauber

Fangen wir mit der wohl wichtigsten Frage überhaupt an. Was ist eigentlich der Sinn des Lebens, des Universums und von überhaupt allem? Auch darauf weiss Google die richtige Antwort: 42. Hierbei handelt es sich allerdings um einen kleinen Scherz, der sich auf das Kultbuch *Per Anhalter durch die Galaxis* des britischen Autors Douglas Adams bezieht. Wenn man den Suchbegriff «tilt» eingibt, neigt sich der gesamte Bildschirminhalt nach schräg rechts. Auch ein Scherz.

Aber wie findet Google Trefferlisten für beantwortbare Fragen? Eigentlich ganz einfach. Indem die Suchmaschine unablässig Heerscharen von sogenannten Crawlers oder Spiders (Spinnen) durch die Weiten des Internets schickt. Das sind kleine Suchroboter in Form von Programmen, die die Inhalte aller öffentlich zugänglichen Webseiten abtasten und ein Verzeichnis der darin enthaltenen Begriffe an riesige Speichercomputer melden. Dort werden diese Begriffe schön durchnummeriert und aufbewahrt. Daraus entsteht eine Trefferliste, wenn ein Benutzer nach einem bestimmten Wort sucht. Versuchen wir es doch gleich mit «Trefferliste». Innerhalb von 0,29 Sekunden spuckt Google eine Liste mit insgesamt rund 47 Millionen Ergebnissen aus. Zuerst – das

kann aber variieren, dazu kommen wir noch – steht die Universität Mainz, dann erscheinen verschiedene Wörterbücher. Das hilft nun nicht unbedingt weiter, wenn man beispielsweise eine Erklärung für dieses Wort sucht. Erste Erkenntnis: Die Antwort von Google ist immer so gut wie die Frage. Wie entsteht nun aber die Reihenfolge, also wieso meint Google, dass die Webseite der Universität Mainz wohl genau das Richtige für den Begriff sei, der hier gesucht wird?

### Der Such-Algorithmus

Das grosse Geheimnis von Google wie natürlich auch von allen anderen Suchmaschinen im Netz ist ein Algorithmus, also eine mathematische Formel, mit der die Suchergebnisse nach Relevanz sortiert werden. Relevanz bedeutet, dass Google erahnen möchte, welche der unzähligen Webseiten, auf denen der gesuchte Begriff vorkommt, wohl die geeignetste für den Sucher ist. Wie dieser Algorithmus genau funktioniert, ist das wohlgehütete Geheimnis von Google. Wäre bekannt, wie die Reihenfolge der Trefferliste hergestellt wird, könnte man sie natürlich manipulieren. Denn das Internet ist ja nicht in erster Linie dafür da, seinen Benutzern Fragen zu beantworten, sondern es ist

ein immer wichtiger werdender Vertriebskanal für Firmen.

Nehmen wir das naheliegende Beispiel, dass ein Student oder ein Schüler einen Computer kaufen will. Also gibt er den Suchbegriff «Discount Computer kaufen» ein und beschränkt die Suche vielleicht auf die Schweiz. Und schwups erscheint die Trefferliste, auf der – wir wollen hier ja keine Werbung machen – diverse Computerfirmen ihre Dienste anbieten. Steht nun zuoberst der günstigste, beste oder zumindest am besten sortierte Shop? Jein. Zuerst und in der Spalte rechts erscheinen bezahlte Inserate. Aber dann kommt die Liste, und da sollte Google doch nach geheimnisvollen, aber korrekten Kriterien die für den Käufer eines Discount-Computers geeignetsten Läden aufgelistet haben? Im Prinzip ja, wenn da die Suchmaschinenoptimierung nicht wäre.

### Ständiger Kampf

Vor allem für Anbieter, die nur einen Internet-Shop haben, ist es natürlich lebenswichtig, dass potenzielle Käufer auf ihr Angebot aufmerksam werden. Es ist eine Erfahrungstatsache, dass kaum jemand mehr als die ersten 10 Treffer genauer anschaut. Also bieten Internet-Agenturen unter anderem die Dienstleistung an, eine sogenannte Suchma-



Das Sortieren der Trefferliste basiert auf einem Algorithmus. Wie dieser funktioniert, ist das wohlbehütete Geheimnis von Google.

schinenoptimierung durchzuführen. Ohne hier technisch-langweilig zu werden, geht es dabei darum, mit den geeigneten Massnahmen dafür zu sorgen, dass bei der Eingabe von bestimmten Suchbegriffen, im Fachjargon Keywords, die so optimierte Firma möglichst weit oben in dieser Liste erscheint. Wie das genau gemacht wird, ist wiederum Geschäftsgeheimnis der Anbieter dieser Dienstleistung. Aber das Resultat ist ja eindeutig überprüfbar. Erscheint ein solcher Shop nach einer gewissen Zeit nicht mindestens in den ersten zehn Treffern, hat's nichts genützt.

## Kein rechtsfreier Raum

Aber das Internet ist nicht nur der grösste Shop der Welt, das Internet ist auch die grösste Ansammlung von vorhandenem Wissen. Informationen jeder Art, die man früher mühselig in Bibliotheken, Nachschlagewerken oder Enzyklopädien zusammensuchen musste, sind im Prinzip nur einen Mausklick entfernt. Damit ist auch das Leben von Lehrpersonen, Schülerinnen und Schülern, Dozierenden und Studierenden entschieden einfacher geworden. Eine Hausarbeit zu jedem beliebigen Thema? Eine wissenschaftliche Auseinandersetzung mit einem Begriff? No problem, kann man da sagen. Aller Wahrschein-

lichkeit nach wird es doch jemanden geben, der genau diese Aufgabenstellung schon bearbeitet und gelöst hat. Und falls die Ergebnisse seiner Anstrengungen digital vorliegen, was heutzutage üblich ist, dann sollte man sie doch auch finden. Und benützen. Also am einfachsten in Form von copy/paste. Spart jede Menge Arbeit. Kann aber jede Menge Ärger einbringen.

Zunächst einmal ist das Internet kein rechtsfreier Raum. Auch hier gilt das Copyright, also das Recht des Urhebers an seinem Werk. Viele Blogger mussten schon erfahren, dass die Verwendung von im Web gefundenen Bildern oder Fotos auf dem eigenen Blog nicht folgenlos bleibt. Denn natürlich gibt es auch dafür Suchmaschinen, die das Netz nach Kopien von urheberrechtlich geschützten Werken absuchen. Werden sie fündig, dann bekommt der Kopist im Ernstfall Post von einer Anwaltskanzlei. Ähnlich verhält es sich auch mit der Verwendung von geistigem Eigentum, welches in Form von Buchstaben im Internet vorhanden ist.

## Plagiate und deren Folgen

Jeder Student muss heute bei Abgabe einer Arbeit versichern, dass er sie selbstständig und ohne fremde Hilfe verfasst hat. Zudem dürfen keine fremden Texte

kopiert und als eigene ausgegeben werden. Merkt doch keiner, sagten sich da schon viele. Das hat allerdings auch schon vielen sogar den Dokortitel gekostet. Denn auch Lehrkräfte haben nachgerüstet; es gibt inzwischen eine Vielzahl von Programmen, die man einsetzen kann, um geklauten Textstellen auf die Spur zu kommen. Oder aber man verwendet die Suchmaschine selbst.

Bei der Eingabe des letzten Satzes dieses Artikels zeigt sich: Es werden keine Ergebnisse gefunden. Test bestanden. Es scheint sich also um einen Originalsatz zu handeln. Wenn diesen aber ein Leser einfach kopieren und verwenden sollte: Vorsicht, Google wird's finden.

**Philipp Sauber** ist Inhaber der Webagentur INM AG ([www.inm.ch](http://www.inm.ch)) und Experte auf den Gebieten Online-Marketing und Weblösungen. Sein Wissen gibt er zusätzlich als Dozent an diversen Weiterbildungsinstituten weiter. [philipp.sauber@inm.ch](mailto:philipp.sauber@inm.ch)