

Suche im Internet, Google und Co.

Sagenhafte Suchmaschinen

Worldwidewebsize.com schätzt die aktuelle Größe des sichtbaren Internets auf 23 Milliarden Seiten. Abgeheftet ergäbe das einen Ordner so breit wie von Bozen bis Madrid. Dazu kommen unzählbare weitere unverlinkte Inhalte und dynamische Datenbankinformationen, die von Suchmaschinen nicht erfasst werden. Wie finden diese Suchmaschinen eigentlich all diese Seiten? Und wie sucht man damit?

VON FRITZ JÖRN *

Tim Berners-Lee, der Erfinder des WWW, führte am Cern 1992 eine Liste der Web-Adressen – 26 Stück. 1993 baute Matthew Gray am MIT seinen „World Wide Web Wanderer“. 1994 entstand schon der erste Volltext-Web-„Krabbler“ namens WebCrawler, und seit 1998 dominiert Larry Page and Sergey Brins Google die Suche.

Das Suchprogramm durchsucht mit affenartiger Geschwindigkeit und mit abertausenden von Maschinen laufend das Internet, von Link zu Link. Ich stelle mir das wie ein Labyrinth vor: Da gibt es auch den Trick, mit einer Hand immer an der Wand entlang zu gehen; so kommt man zwar nicht besonders schnell, aber ganz gewiss heraus. Ist Eingang gleich Ausgang, so berührt man jede Wand (außer Inseln). Ähnlich bekommen Suchmaschinen jede mit einem Link eingebundene Seite zu sehen, nicht verlinkte Inseln aber nicht. Sie merken sich in einer „umgekehrten Liste“ zu allen gefundenen Begriffen die Adresse der Seite: Sie „indexieren“ in eine invertierte Liste. Bei der tatsächlichen Suchabfrage wird dann nicht mehr gesucht, sondern in unter einer halben Sekunde der Inhalt des Indexes wiedergegeben.

Noch ein paar Suchtipps, hauptsächlich für Google. Die Beschäftigung mit der Google-Suche lohnt sich – auch, weil man sie doch häufig verwendet. Zunächst stellt man sich unter „Einstellungen“ die Sprache und mehr ein. Ganz praktisch: Setzt man bei der Suche Begriffe in gerade Anführungszeichen, etwa "Suchtipps, hauptsächlich für Google", so wird nur nach Stellen gesucht, die genau diesen Text enthalten. Auf Groß- und Kleinschreibung, Kommas usw. wird dabei nicht geachtet. Damit kann man sich bei großen Internet-publizierten Dokumenten das Herummailen sparen: Man greift sich ein Stück Text aus dem Dokument ab, etwa "Gebäude Räumlichkeiten oder Gruppen", und lässt Google

Mit einer Suchmaschine lässt sich – fast – alles finden.

suchen, ob der Text schon irgendwo im Web steht. Danach reicht in einer E-Mail die Internet-Adresse als Verweis statt eines riesigen Dateianhangs.

Ein Stern im Suchtext steht als Platzhalter für Beliebiges, bitte probieren. Das logische Oder wird mit einem senkrechten Strich (|, altGr<) oder OR ausgelöst. Die Tilde ~ sucht nach Ähnlichem. Ein Minus meidet den unmittelbar folgenden Begriff. „Cache:http://usw.“ zeigt die zuletzt gespeicherte Seitenkopie, dazu das Besuchsdatum von Google. „Link:“ gibt an, wer alles zur Seite verlinkt. Schließlich beschränkt „filetype:“ die Suche auf bestimmte Dateiformate, praktisch etwa „filetype:jpg“, besonders am eigenen Rechner. Gezielt nur auf einer bestimmten Website sucht man mit „site:“ im Suchfeld, etwa „site:bz.it“ und dann dem Suchbegriff. Und als schneller kleiner Rechner funktioniert das Google-Eingabefeld auch. Unter „Mehr“ versucht sich Google sogar als Übersetzer. Noch mehr steht englisch auf <http://jweb-net.net/advancedgooglesearch.html>.

Bilder findet man mit der extra Bildersuche, wobei man freilich davon abhängig ist, mit welchem Namen und in welchem Kontext ein Bild ins Netz ge-

stellt worden ist. Da erscheint schon einmal auf der Suche nach einem „Hof“ die Jennifer gleichen Namens – auch schön.

Die Suche in Google-eingescannten Büchern – eine Fundgrube! – passiert nicht automatisch mit, man muss über „Mehr“ zu „Bücher“ verzweigen. Aber dann bringt uns „Ritterbahn“ zur Zeitschrift „Schlern“ von 1972, Seite 192, und zu einem Artikel von Walther Amonn: „Abschied von der Rittner Bahn“. Peinlich nur, dass wir das der University of Michigan verdanken müssen, und schade, dass es den Inhalt dann nicht zu lesen gibt.

Schließlich bietet uns Google mit dem Gratisprogramm „Google Desktop“ an, den eigenen PC zu indexieren (Vista soll das selbst können, Copernic, Yahoo und andere auch).

Bei Google-Desktop bestimmt man unter „Desktop-Einstellungen“, wenn andere PCs und Netzlaufwerke mit indexiert werden sollen oder umgekehrt bestimmte Laufwerke – etwa das Backup – ausgenommen bleiben. Für ausgewählte Dateien und meine E-Mails am PC bevorzuge ich zum Suchen eine alte, kostenlose Version 1.3.0 von „Lookout for Outlook“ (bis Outlook 2003, noch von

www.majorgeeks.com/Lookout_d4808.html zu holen).

Metasuchmaschinen lassen sich von mehreren Suchmaschinen Ergebnisse liefern. Spezialsuchen bedienen Sonderinteressen, etwa Google-Blogsuche oder Google Maps auf der Landkarte. Keypict Photo Search indexiert für Google-Desktop versteckte XMP- und IPTC-Eintragen aus Fotos. Eurobuch.com sucht alte Bücher, Ortssuchen wie Maxmind.com erraten aus der IP-Nummer einen Ort. Musik, Videos – alles lässt sich im Netz suchen.

Schließlich: Nett zu sehen, wie sich ein Webauftritt historisch entwickelt hat. Dazu adressiert man Archive.org und gibt die Webadresse ein. Gespeichert sind 85 Milliarden vergangene Web-Seiten seit 1996. Das alles kostet nichts. Das Internet ist ein ethisches Phänomen. Viel Freude. **W**



* Fritz Jörn ist freier Journalist in Bonn



Foto: Schierenbeck/tmm