

BLACKBOX SUCHMASCHINE

POLITIK FÜR NEUE MEDIEN

INTERVIEW MIT MARCEL MACHILL UND WOLFGANG SANDER-BEUERMANN

KAI LEHMANN

Google & Co. sind die zentralen Marktplätze im Internet; ohne sie würden sich nur wenige Menschen im Web zurechtfinden. Über 90 Prozent aller Internet-Nutzer greifen deshalb auf Suchmaschinen zurück [1] – nicht immer als kundige Verbraucher, denn sowohl das WWW als auch die Suchmaschinen haben ihre Untiefen.

Es beginnt beim Netz der Hyperlinks: Die Zahl der weltweit vorhandenen Webseiten ist technisch nicht zu bestimmen. Die Universität Bielefeld schätzt die Größe des frei zugänglichen WWW Anfang 2005 auf 10-15 Milliarden Seiten [2]; Informationen in Datenbanken, auf geschlossenen und dynamischen Webseiten sowie die mit Webseiten verknüpften Dokumente hier nicht mit eingerechnet. Bei dieser Zahl wird klar, wie wichtig Suchmaschinen sind, um Informationen zu finden. Doch erfassen diese bei weitem nicht alles: Zwar verzeichnete Google im Januar 2005 laut eigener Aussage rund 8 Milliarden Webseiten, doch dürfte ein erheblicher Teil des WWW fehlen. Für Deutschland errechnete Andreas Pothe in einer Untersuchung für das Regionale Rechenzentrum für Niedersachsen, dass von den geschätzten 320 Millionen deutschen Webseiten nur 60 Prozent beim Marktführer Google zu finden seien [3]. Schlusslicht Fireball kennt gerade einmal jede zweite deutsche Domain – und entsprechend weniger Webseiten.

Suchmaschinen sind aber nicht nur lückenhaft, sondern auch Torwächter in einem doppelten Sinne: Nur wer es auf die vorderen Plätze eines Suchergebnisses schafft, ist für Nutzer wirklich sichtbar. Spätestens auf der dritten Seite der Suchtreffer schaut kaum noch ein Wissenshungriger nach. Und nur der Nutzer, der die Suchmaschinen kompetent bedienen kann, entlockt ihr alle gewünschten Geheimnisse – von Zensur, blinden Flecken und technischen Unzulänglichkeiten einmal abgesehen. Bei den vielen Mängeln, die die Informationssuche mit sich bringt, wiegt es umso schwerer, dass 70 bis 90 Prozent der Datenreisenden den Marktführer Google – die meisten von ihnen ausschließlich – nutzen. Damit ist der Marktanteil des einstigen Insidertipps erdrückend. Studien zeigen,

dass die Info-Hungrigen auch beim Suchen selbst einseitig und intuitiv herangehen – mit allen Konsequenzen für die gefundenen Ergebnisse: Eine Studie der Universität Karlsruhe hat bei 6.000 Befragten – meist erfahrenen – Nutzern herausgefunden, dass diese zu 70 Prozent nur die ersten fünf Treffer einer Suchmaschine berücksichtigen [4]. Da fällt das so genannte Spamming von Suchergebnissen besonders ins Gewicht: Dubiose Anbieter versuchen mit allen Mitteln die Websites ihrer Kunden auf die vorderen Plätze der Ergebnisse zu katapultieren – unabhängig vom gesuchten Thema. Auch der Jugendschutz wird gern bemüht, um die mangelnde Qualität der Suchmaschinen zu beklagen: Neben pornografischen Angeboten werden immer wieder rechtsradikale Webseiten als Beispiel angeführt, die insbesondere für Jugendliche gefährlich seien (vgl. dazu kritisch [5]). Aus welchem Blickwinkel Suchmaschinen auch betrachtet werden: Sie spielen in der vom Internet geprägten Gesellschaft eine zentrale Rolle – mit allen positiven und negativen Folgen. Unterschiedliche Vorschläge wurden gemacht, um zu klären, auf welche Weise die Suchmaschinenbetreiber ihrer Verantwortung gerecht werden, wie Staat oder Netzgemeinschaft reagieren sollten.

Prof. Dr. Marcel Machill von der Universität Leipzig hat im Rahmen eines Projekts für die Bertelsmann Stiftung an der Entwicklung eines »Code of Conduct« (CoC) mitgewirkt. Der CoC legt den Betreibern von Suchmaschinen nahe, sich einer Selbstverpflichtung zu unterziehen. Einen anderen Ansatz verfolgt Dr. Wolfgang Sander-Beuermann vom Rechenzentrum Hannover, der 2004 einen Verein zur Förderung der Suchmaschinen-Technologie und für den freien Wissenszugang [6] mitgegründet hat. Beide erläutern im Folgenden, welches Verhältnis Gesellschaft und Suchmaschinen zueinander bekommen sollten.

Herr Machill, der Bundestags-Ausschuss »Neue Medien« bezeichnet die Situation am Suchmaschinenmarkt als unbefriedigend. Wie ernst ist es mit der Monopolstellung von Google?

Der Marktanteil von Google am deutschen Suchmaschinenmarkt liegt je nach Erhebungsmethode zwischen 69 und 83 Prozent. Der größte Mitbewerber Yahoo! erreicht in Deutschland gerade einmal knapp zweistellige Marktanteile. Insofern kann man bei Google durchaus von einem Monopol reden – allerdings mit Einschränkungen, denn niemand ist gezwungen, für eine Suche auf Google zurückzugreifen. Die Nutzer verzichten leider allzu oft darauf, Zweit- oder Drittquellen zu befragen. Alternativen gäbe es genug. Googles technologischer Vorsprung schmilzt aber zusammen und die anderen Suchmaschinen werden aufholen. Dies erklärt die vielen Aktivitäten von Google abseits der Websuche; es geht

um die Bindung von Nutzern an die Marke. Auf lange Sicht werden wir uns in Deutschland auf Verhältnisse ähnlich wie in den USA einstellen müssen, wo sich einige große Suchmaschinen den Markt teilen.

Welche Stolpersteine birgt der Umgang mit Suchmaschinen?

Zuerst einmal die Annahme, mit einer Suchmaschine könne man sich einen umfassenden Überblick verschaffen. Google hat mittlerweile über acht Milliarden Internetseiten in seinem Index, aber auch dies ist nur ein Bruchteil des gesamten Netzes. Eine annähernd brauchbare Übersicht über ein Thema erhält nur, wer mehrere Suchmaschinen parallel nutzt. Viele Informationen sind im Netz auch gar nicht auffindbar, sondern machen immer noch den Gang in Bibliotheken notwendig.

Den Nutzern fehlt das Wissen für den Umgang mit Suchmaschinen. Viele arbeiten nur mit wenigen, oft sehr allgemein gehaltenen Schlüsselwörtern. Das führt dazu, dass die Suchmaschine eine Masse von Suchergebnissen präsentiert, von denen der Großteil durch Werbung oder Spam nicht relevant ist. Wie man eine Suche sinnvoll eingrenzen kann, haben viele Menschen nicht gelernt. Hinzu kommt, dass sich in den Ergebnislisten Ergebnisse tummeln, die dort überhaupt nicht hingehören, z.B. Werbung oder Spam.

Wie setzt der von Ihnen erarbeitete Code of Conduct (CoC) hier an?

Die Nutzer brauchen vor allem Transparenz bei den Ergebnissen. Es muss klar sein, wie Suchergebnisse zu Stande kommen, etwa wie die Suchmaschine gewichtet und ob ein Link werbefinanziert ist oder aber aus der freien Websuche stammt. Auch ist der Aspekt des Rückkanals zur Suchmaschine von großer Bedeutung: Die Nutzer brauchen einfach Ansprechpartner, bspw. eine Hotline, wenn sie Probleme haben.

Des Weiteren müssen Bewertungsmechanismen für die Suchmaschinen selbst geschaffen werden. Wir hatten hier ein Gütesiegel und eine Freiwillige Selbstkontrolle Suchmaschinen (FSS) angeregt.

Erleichtern transparente Bewertungsverfahren bei Suchmaschinen den Spammern nicht die Arbeit?

Im Gegenteil – mehr Transparenz erschwert Spammern die Arbeit. Notorische Spammer verfügen auch heute schon über umfangreiches Wissen über die Funktionsweise von Suchmaschinen. Es gibt eine ganze Branche, die Search-Engine-Optimizer (SEO), die im Versuch-und-Irrtum-Verfahren die Funktionsweise der einzelnen Suchmaschinen bis ins Detail auskundschaften, um dann ihre Kunden gezielt in den Ergebnislisten zu pushen. Nun sind SEOs nicht automatisch Spammer, aber sie liefern den echten Spammern oftmals das notwendige Wissen. Wenn

nun aber jeder normale Nutzer durch mehr Transparenz weiß, wie die Suchalgorithmen funktionieren, so kann er seine eigene Internetseite entsprechend gestalten, dass sie auch gefunden wird. Gleichzeitig können die Suchmaschinen viel restriktiver gegen Spam vorgehen. Die Schaffung von Transparenz muss einhergehen mit klaren Regeln, wo Spam beginnt – und dann können sie die Spammer auch sanktionieren, sprich aus dem Index löschen. Diese zwei Seiten der Medaille stehen explizit im CoC.

Einige Suchmaschinen haben sich dem CoC schon verpflichtet. Mehr Transparenz ist aber nicht erkennbar. Wie beurteilen Sie die Umsetzung und Effizienz der CoC-Regeln?

Die Regeln des CoC bedürfen einer neutralen Kontrollinstanz, sprich einer Institution à la FSS. Bisher fehlt diese Instanz jedoch, weshalb die Ergebnisse noch nicht so sind, wie sie sein könnten. Es fehlen noch die Anreize, beispielsweise ein marktwirksames Gütesiegel, damit sich diese Regeln im hart umkämpften Markt stärker durchsetzen werden. Hier könnte die (deutsche) Medienpolitik einiges tun, wenn sie nur wollte.

Studien belegen regelmäßig, dass das unreflektierte Nutzerverhalten erheblich zur Situation beiträgt. Wie ist hier Abhilfe zu schaffen?

Mediennutzung muss bereits in den Schulen stärker verankert werden, nicht nur das Internet betreffend, sondern auch im Hinblick auf Bücher, Zeitschriften und den Rundfunk. Wir leben in einer Wissensgesellschaft, und da gilt mehr denn je: »Wissen ist wissen, wo's steht«. Hierfür brauchen wir kein neues Fach, aber eine leichte Korrektur aller Fächer – wir müssen Schülern vernünftiges Recherchieren beibringen.

Wie verhält sich Google als führende Suchmaschine zum CoC?

Aus der Technologieführerschaft von Google ergibt sich automatisch, dass man sich – vor allem im Hinblick auf die Transparenz des Ranking – bedeckt hält. Auch kollidiert die Forderung nach Datensparsamkeit teilweise mit Googles Geschäftsmodellen, bspw. AdSense. Hier besteht großer Nachholbedarf. Andererseits ist Google bei der Kennzeichnung von bezahlten Links, sprich Werbung, sehr vorbildlich zu nennen.

Der CoC plädiert dafür, bestimmtes Wissen auszublenden, indem die Suchergebnisse gefiltert werden. Wie steht es um das Recht auf einen freien Informationszugang?

Es geht vor allem darum, nach geltendem Recht als illegal eingestufte Inhalte auszublenden, bspw. Nazi-Seiten, die Auschwitz leugnen oder Seiten mit Kinderpornographie, also Informationen, auf die sie auch in der Offline-Welt kein Recht auf Informationszugang besitzen. In puncto

Jugendschutz muss man fragen, was mehr wiegt: das Recht der Jugend auf eine ungestörte Kindheit oder das der Erwachsenen auf einfache, möglichst kostenlose Triebbefriedigung? Ich plädiere für eine publizistische Verantwortung der Suchmaschinenbetreiber, nicht für Zensur.

Wer entscheidet, welches Wissen für welche Nutzer gesperrt wird?

Die Entscheidung über den Ausschluss bestimmter Seiten sollte eine politisch wie wirtschaftlich unabhängige Kontrollinstanz wie die FSS ausüben. Es müssen im Vorfeld klare Richtlinien gesetzt werden, welche Inhalte ausgeschlossen werden können, und es muss explizit das Recht geben, dass man gegen diese Entscheidung vor Gericht ziehen kann und auf Wiederaufnahme der Seite in den aktiven Suchmaschinenindex klagt.

Lässt sich der Zugang zu Wissen im Netz effizient einschränken?

Wer technisch versiert ist, kann sich immer Zugang zu Informationen beschaffen. Sie können bspw. die deutschen Google-Seiten umgehen, indem sie einfach auf google.com/ncr wechseln, dann steht ihnen nämlich der gesamte US-Index für ihre Suche offen. Das ist aber kein Argument dafür, den Jugendschutz *nicht* zu verbessern. Sonst müsste man nämlich mit ähnlicher Argumentation jegliche Strafverfolgung einstellen, nur weil es uns nicht gelingt, eine verbrechensfreie Gesellschaft zu schaffen.

Welches Fazit ziehen Sie für den CoC im Jahre 2005? Dient er nicht viel zu sehr als Feigenblatt für die teilnehmenden Suchmaschinen?

Die Idee des CoC war von Anfang an mit der Idee einer unabhängigen Kontrollstelle verbunden, die dessen Einhaltung kontrolliert. Insofern ist der derzeitige Stand nicht befriedigend. So lange es keine FSS gibt, wird auch die Umsetzung des CoC in einigen Fällen nur Lippenbekenntnis bleiben. Hauptproblem bei der Schaffung einer Freiwilligen Selbstkontrolle Suchmaschinen ist wohl die Finanzierung.

Wo sehen Sie die Informationskultur im Jahre 2010?

Die Nutzer werden sich viele Freiräume im Netz, die heute kommerzialisiert sind, zurückerobern. Der Firefox-Browser ist nur ein Beispiel dafür. Die Open-Source-Bewegung wird ihr Pendant auch in der Informationswelt finden. Schon heute wird eifrig an einer Open-Source-Suchmaschine [7] gearbeitet. Das revolutionäre an Google 1998 war ja, dass die riesigen Datenmengen auf einem handelsüblichen Rechner verwertet werden konnten. So wie Tausende Nutzer am Seti@Home-Projekt teilnehmen, ist auch eine dezentrale Suchmaschine denkbar, der die Nutzer dann die freie Rechenzeit ihres Computers spendieren können.

Herr Sander-Beuermann, wie bewerten Sie die Stellung von Google im Internet? Ist der freie Zugang zu Wissen gefährdet?

Der freie Zugang zum digitalen Weltwissen ist durch jedes De-facto-Wissensmonopol gefährdet. Eine solche Situation ist vergleichbar derjenigen, bei der es in den Print-Medien nur noch eine einzige Zeitung oder nur noch eine Bibliothek auf der Welt gäbe. Dabei ist es gleichgültig, wie der De-facto-Monopolist gerade heißen mag. Derzeit ist es Google, aber die Situation würde durchaus nicht besser, wenn bspw. der Erzkonkurrent Microsoft diese Position übernehme.

Der Code of Conduct versucht, Suchmaschinen zur Eigeninitiative anzuregen. Für wie erfolgversprechend halten sie diesen Ansatz?

Dieser Ansatz geht in eine andere Richtung: Es geht dabei nicht um Monopole, sondern um eine freiwillige Selbstverpflichtung von Suchmaschinenbetreibern hinsichtlich der findbaren Inhalte; für Deutschland spielt das Jugendschutzgesetz bspw. eine wesentliche Rolle. So etwas ist bei der Globalität des Internet und der Regionalität der unterschiedlichen Rechtsnormen sicherlich noch schwieriger – der vorgeschlagene Code of Conduct der Bertelsmann Stiftung blieb erfolglos. Die FSM (Freiwilligen Selbstkontrolle Multimedia) versucht jetzt in dieser Frage zu praktikableren Lösungen zu kommen.

Oft werden die intransparenten Bewertungskriterien der Suchmaschinen kritisiert – allen voran »PageRanking« von Google, weil es auf Masse statt Klasse setzt. Welche Vor- und Nachteile sehen sie?

Das PageRank-Verfahren von Google war vor Jahren ein bahnbrechender Erfolg. Es ist im Prinzip auch heute noch nicht völlig verkehrt, selbst wenn es mittlerweile eine Art »Sport« von Suchmaschinen-Optimierern geworden ist, das Verfahren auszutricksen. Das Problem liegt woanders: Wenn es fast nur noch eine einzige Suchmaschine gibt, dann entscheidet allein deren Ranking-Verfahren über Sein oder Nicht-Sein im Web. Wenn es hingegen eine Vielfalt von Informationsanbietern gibt, dann gibt es auch eine Vielzahl an Ranking-Verfahren (die auch nicht alle gleichzeitig technisch austricksbar sind). Das Ziel muss also sein, hier wieder eine Vielfalt herzustellen.

Verglichen mit der großen Kritik an Google, gibt es in Deutschland kaum Gegenwind für den Marktführer. Traut sich niemand?

Ich denke, es ist eher so, dass die meisten Nutzer hier noch gar kein Problembewusstsein entwickelt haben: »Ich nehme immer Google, die sind doch gut«, so das häufige Credo. Dass sie vielleicht doch nicht so gut sind, zeigen dann erst Berichte wie »Google zugemüllt« (c't, Heft 20,

2003) oder Pressemeldungen wie »Suchmaschinenbetreiber ist Komplize der Zensoren in Peking« [8]. Wenn sich jemand tatsächlich nicht trauen sollte, gegen Google Initiative zu ergreifen, dann müsste er schon die »tiefe Einsicht verinnerlicht« haben, dass dann Google eventuell seine Existenz im WWW auslöschen könnte. Aber diese Einsicht ist noch so selten, dass dies kaum der Grund sein wird.

Wie können Alternativen zu Wissensmonopolen aussehen?

Alternativen müssen möglichst unabhängig und frei, und allein durch ihre Organisation und Struktur prinzipiell kaum monopolisierbar sein. Dazu sehe ich drei Möglichkeiten:

1. *Die öffentlich-rechtliche Suchmaschine:* Unabhängigkeit könnte einerseits der Staat durch eine öffentlich-rechtliche Suchmaschine etablieren, genauso wie er das seit langen Jahren in den konventionellen Medien tut – es ist eigentlich nahe liegend, dies für die Neuen Medien zu übernehmen. Wesentlich ist dabei jedoch, dass dann nicht »der Staat selbst« wiederum Monopolist wird – also müsste die Unabhängigkeit einer öffentlich-rechtlichen Suchmaschine durch Gremien und Beiräte abgesichert werden, was etliche Folgeprobleme aufwirft.

2. *Peer-to-Peer:* Auf der anderen Seite eines Spektrums möglicher Alternativen steht das genaue Gegenteil: ein unstrukturiertes Gebilde aus lose kooperierenden kleinen Suchmaschinen, die sich zu einem großen Datenraum zusammenschließen. Technisch kann so eine Lösung durch eine Peer-to-Peer (P2P)-Software realisiert werden; hierzu gibt es auch bereits einen funktionierenden Ansatz, das YACY-Projekt von Michael Christen [9]. Der Vorteil einer solchen unstrukturiert verteilten Suchmaschinen-Infrastruktur ist seine inhärente Nicht-Monopolisierbarkeit: Niemand kann ein solches Gebilde in seiner Gesamtheit kontrollieren. Der Nachteil ist, dass diese strukturell in ihren Wegen nicht vorab definierten Datenflüsse nicht unbedingt effektiv, also meist zeitaufwändig sind – der Nutzer muss länger auf das Ergebnis warten. Die von Google gewohnten Antwortzeiten unter einer Sekunde sind illusorisch.

3. *Kombinationen aus beiden Ansätzen.* Sie stehen zwischen den beiden genannten Enden des Spektrums. Am sinnvollsten erscheint mir, ein Verbund aus Hunderten von kleinen Suchmaschinen (Mini-Suchern) teilt den zu erfassenden Datenraum unter sich auf; diese Absprache kann in der Anfangsphase tatsächlich durch verbale Kommunikation geschehen, sollte später aber automatisiert zwischen den beteiligten Rechnern ablaufen. Betreiber solcher Mini-Sucher können nun einerseits Privatpersonen oder Vereine sein, andererseits aber Organisationen des »öffentlich-rechtlichen Raumes«, wie Bibliotheken und Universitäten. Letztere bieten Verlässlichkeit staatlicher Institutionen, sind aber in Grenzen

durchaus unabhängig. Private Betreiber können ihrerseits auch Peer-to-Peer-Datenquellen einspeisen und so die weitere Unabhängigkeit sichern. Um diesen Verbund für den Nutzer zusammenzuführen, und einen einfachen Einstieg zu ermöglichen, können Meta-Sucher als zentrale Einstiegspunkte dienen. Dreh- und Angelpunkt ist die Organisation und das kooperative Zusammenwirken aller Teilnehmer: Jeder kann mitmachen, muss sich aber in Schnittstellen einpassen und kann bei Missbrauch ausgeschlossen werden.

Was ist das Besondere an dem Projekt Nutch?

Nutch ist eines von mehreren Projekten, die freie Suchmaschinensoftware auf Open-Source-Basis entwickeln – neben »Mngosearch« und »Aspseek«, die für bis zu ein paar Millionen Webseiten geeignet sind. In den USA wurde im Umfeld von www.archive.org »Heritrix« entwickelt, mit bereits deutlich höheren Ambitionen. Allein Nutch aber hat das hohe Ziel, eine Alternative zu Google zu werden und viele Milliarden Webseiten zu erfassen. Bis dahin ist es noch ein weiter Weg. Ich halte es für unrealistisch, diesen Weg gradlinig verwirklichen zu können: Allein die zu bewegendenden Datenströme verlangen immense Ressourcen. Aber Nutch könnte in einem »Konzert von vielen die erste Geige spielen«.

Wo sehen Sie die digitale Informationskultur im Jahre 2010?

Wenn ich einen pessimistischen Tag habe, dann sehe ich das, was an Informationskultur noch übrig bleibt, in der Hand von ein oder zwei globalen Konzernen, die damit unser Denken und unser Weltbild beliebig formen können. Meist aber sehe ich die Welt optimistisch und als eine kooperative Wissensinfrastruktur, an der jeder teilnehmen kann – als Konsument, Lieferant und Mitgestalter. Aber Optimismus allein genügt nicht, um Visionen zu realisieren. Daher habe ich im Juli 2004 den »Gemeinnützigen Verein zur Förderung der Suchmaschinentechnologie und des freien Wissenszugangs« initiiert, kurz SuMa-eV [6].

Digitale Verweise

- [1] www.politik-digital.de/text/econsumer/verbraucherschutz/guetesiegel.shtml
- [2] www.ub.uni-bielefeld.de/biblio/search/help/statistik.htm
- [3] www.heise.de/newsticker/meldung/53568
- [4] <http://marketing.wiwi.uni-karlsruhe.de/institut/mitarbeiter/suchverhalten.jsp>
- [5] <http://odem.org/informationsfreiheit/zahlen-naziseiten.html>
- [6] www.suma-ev.de
- [7] www.nutch.org
- [8] <http://de.internet.com/index.php?id=2032510>
- [9] <http://suma-lab.de:8080>