

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/256473910>

# Search Engine Bias

Chapter · January 2011

CITATIONS  
0

READS  
447

1 author:



[Karsten Weber](#)  
Ostbayerische Technische Hochschule  
190 PUBLICATIONS 207 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Identifikation von Fahrttypen [View project](#)



Der Einfluss von Fach- und Onlineberatung bei der Behandlung von MS – eine Längsschnittuntersuchung. [View project](#)

# Search Engine Bias

Karsten WEBER

*Professor für Allgemeine Technikwissenschaften  
BTU Cottbus, Postfach 101344, 03013 Cottbus  
Karsten.Weber@tu-cottbus.de*

**Abstract.** Der Name ‚Google‘ wird heute synonym für ‚Suchmaschine‘ genutzt; das Verb ‚googeln‘ hat es zu einem Eintrag im Duden gebracht: *Google* ist allgegenwärtig. Dies bestätigen Statistiken, die aufweisen, dass *Google* in vielen Ländern andere Suchmaschinen völlig marginalisiert hat. Außerdem zeigen Untersuchungen, dass sich viele Nutzerinnen und Nutzer auf die Ergebnisse einer einzigen Suchmaschine – oft eben *Google* – verlassen. Daher stellt sich die Frage, ob die tatsächliche Qualität der Suchergebnisse das Vertrauen in die angenommene Qualität einer Suchmaschine rechtfertigen kann. Nach einleitenden Bemerkungen zum Suchmaschinenmarkt und einer kurzen Beschreibung der generellen Funktionsweise heute verbreiteter Suchmaschinen wird evaluiert, welche Faktoren die Ergebnisse einer Suchmaschine determinieren. Dabei wird sich zeigen, dass Suchergebnisse durch vielfältige Faktoren beeinflusst werden, so bspw. durch das Nutzerverhalten auf verschiedenen Ebenen, durch die verwendeten Indexierungs- und Suchalgorithmen oder durch die Inhalte des Netzes selbst. Diese und weitere Faktoren führen zu Verzerrungen der Suchergebnisse gegenüber einem unterstellten Idealergebnis. Es zeigt sich, dass es in vielen Situationen sinnvoll ist, sich nicht nur auf eine Suchmaschine zu verlassen bzw. den Ergebnissen einer Suchmaschine nicht blind zu vertrauen. So ist im wissenschaftlichen oder journalistischen Kontext dringend die Nutzung von bereichsspezifischen Suchmaschinen und anderer Rechercheformen angeraten.

**Keywords.** Suchmaschine, Suchergebnis, Verzerrung, Fehler, Bias.

## 1. Suchmaschinen: Die 1000 Augen des Informationszeitalters?

Schon seit geraumer Zeit ist das Internet ein Thema nicht nur für Nerds, Geeks und Hacker, sondern Gegenstand teils heftig geführter öffentlicher Debatten. Standen zu Beginn dieser Erörterungen in erster Linie die ökonomischen Chancen des Internet im Zentrum der Überlegungen und danach die Frage, warum es zu der berühmt-berüchtigten Dotcom-Blase der späten 1990er Jahre kommen konnte, folgte als Nächstes die Diskussion um die digitale Spaltung bzw. die Gefahr, dass das Internet bspw. in Deutschland nicht ausreichend verbreitet sei – derzeit wird dies im Rahmen der Breitbandinitiative der Bundesregierung erneut thematisiert. Seit einigen Jahren nun rücken die vermeintlichen und tatsächlichen Gefahren des Internet in den Fokus der Öffentlichkeit: Über das Netz verbreitete Kinderpornographie, Extremismus und Terror im Netz, Cyberwar, die Datensammelwut staatlicher Institutionen mithilfe von Onlinedurchsuchungen und Vorratsspeicherung oder die Veränderungen persönlicher Beziehungen durch soziale Netzwerke sind nur einige der zuweilen sehr hitzig debattierten Themen. Derzeit dominiert aber wird die öffentliche Auseinandersetzung durch die Kontroversen um die gigantischen Datenberge, die soziale Netzwerke wie *Facebook* und Suchmaschinen wie

*Google* ansammeln und zu ihren Zwecken speichern, verarbeiten und auswerten. Die Hauptsorge ist, dass private Unternehmen, die zudem nicht den deutschen Gesetzen und auch nicht den europäischen Regulierungsmaßnahmen unterworfen sind, diese Daten gegen die Interessen der Nutzerinnen und Nutzer des Internet verwenden könnten.

Entzündet hatte sich der Konflikt rund um *Google* – andere Suchmaschinen spielen in dieser Debatte keine Rolle – an einem neuen Dienst des Unternehmens: *Google Streetview*. Doch die Kombination eines Online-Kartendienstes mit der Bereitstellung von Straßenansichten deutscher Städte und Gemeinden war nur der vordergründige Auslöser; die tiefer liegende Ursache für das Unbehagen *Google* gegenüber ist darin zu sehen, dass diese Suchmaschine so viel über die Welt und uns zu wissen scheint.<sup>1</sup> Einerseits ist dies ein Vorteil: Wir haben zumindest den Eindruck, dass man alles findet, wenn man *Google* fragt. Andererseits ist dies wiederum auch die Ursache für all die Aversionen, die *Google* derzeit entgegengebracht werden: Diese Suchmaschine scheint einfach alles zu wissen – auch über einzelne Personen; wir fühlen uns in unserer Privat- und Intimsphäre bedroht (vgl. bspw. [1], [2] und [3]).

### 1.1. Suchmaschinenmonopoly

Tatsächlich wird der Name ‚Google‘ inzwischen synonym für ‚Suchmaschine‘ genutzt; das Verb ‚googeln‘ hat es zu einem Eintrag im Duden gebracht: *Google* ist allgegenwärtig. Das kann man auch in Zahlen ausdrücken. *Google* nimmt mit seiner Suchmaschine in vielen Ländern eine marktbeherrschende Position bis hin zu einer faktischen Monopolstellung ein; jene Länder, für die dies nicht gilt, ändern nichts daran, dass *Google* mehr als 60% aller weltweit gestellten Suchanfragen beantwortet (siehe Tabelle 1).<sup>2</sup>

Aus ökonomischer Sicht könnte diese Situation bereits als kritisch angesehen werden, da in den mit entsprechenden Fragen beschäftigten wissenschaftlichen Disziplinen vermutet wird, dass die Entwicklung von Monopolen in der Regel zum Nachteil der Verbraucherinnen und Verbraucher gereicht, Innovationen hemmt und kaum mehr aufzubrechende Abhängigkeiten erzeugt (vgl. [4]).<sup>3</sup> Zudem deuten entsprechende Statistiken zusammen mit Ergebnissen anderer Studien (bspw. [5]) darauf hin, dass sich die Nutzerinnen und Nutzer von Suchmaschinen oftmals ausschließlich auf eine Suchmaschine und deren Suchergebnisse verlassen. Zunehmend wird diese Einseitigkeit sogar in professionellen Kontexten beobachtet, bspw. im Rahmen der journalistischen Recherche, sodass hier inzwischen von „Googleisierung“ des Journalismus (vgl. [6]) gesprochen wird.

<sup>1</sup> Zur generellen Bedeutung von Suchmaschinen für Gesellschaften siehe bspw. [7] und [8].

<sup>2</sup> Glaubt man anderen Statistiken, liegt diese Zahl sogar noch höher bei ca. 80%, siehe dazu <<http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4>>, zuletzt besucht am 17.12.2010.

<sup>3</sup> Davon unabhängig zu diskutieren ist die Frage, ob der Suchmaschinenmarkt dazu tendiert, ein natürliches Monopol auszubilden, weil ein Anbieter diesen Markt stets günstiger bedienen kann, als dies zwei oder mehrere Anbieter könnten. Wenn diese Frage mit ja beantwortet werden müsste, hieße dies bspw., dass dadurch ein weitaus höherer Regulierungsbedarf entstünde als dies für einen Markt mit Wettbewerb der Fall ist. Denn der Monopolanbieter einer Suchmaschine besäße ja nicht nur eine immense Marktmacht, sondern würde letztlich den Zugang zu den Ressourcen des Internet weitgehend allein kontrollieren. Doch wie dies politisch zu gestalten wäre kann hier nicht weiter untersucht werden; zudem stellt sich das Problem, dass nationalstaatliche Gesetze für die Regulierung weltweit operierender Suchmaschinen untauglich wären.

**Tabelle 1:** Marktanteile von Suchmaschinen in ausgewählten Ländern<sup>4</sup>

	Brasilien	Deutschland	Frankreich	UK	Indien	Japan	Niederlande	Russland	Südkorea	VR China	USA
AOL											4
Ask				2							2
Baidu										61	
Bing	1		3	2		2					11
Daum									15		
Google	97	89	90	91	82	38	95	31	1	27	63
Ilse							1				
Naver									69		
Orange			2								
Rakuten						2					
Rambler								10			
Search Mail								7			
Sogou										3	
T-Online		3									
Yahoo	1	1	3	3	10	51			5	2	16
Yandex								47			
Andere	1	7	2	3	8	7	4	5	10	7	4

### 1.2. Wie mächtig sind Suchmaschinen wirklich?

Wir stehen also vor der paradoxen Situation, dass sich einerseits *Google* in vielen Ländern zu der Suchmaschine schlechthin entwickelt hat und ihr gleichzeitig zunehmendes Misstrauen entgegengebracht wird. In beiden Fällen aber wird vorausgesetzt, dass *Google* tatsächlich alles weiß und alles findet – worauf immer sich dieses ‚alles‘ auch beziehen mag. Unabhängig von der oftmals sehr aufgeregten und zuweilen polemisch geführten Debatte um die Gefährdung der Privatsphäre und des Datenschutzes durch *Google* liegt daher die Frage nahe, ob denn die Hoffnungen bzw. Befürchtungen hinsichtlich der Informationsomnipotenz von *Google* berechtigt sind. Allgemeiner ausgedrückt: Sind Suchmaschinen tatsächlich jene mächtigen Werkzeuge zum Auffinden aller Informationen, die im Internet verfügbar sind?

Der Titel des vorliegenden Beitrags deutet ja bereits an, dass die Antwort auf die eben gestellte Frage zumindest nicht uneingeschränkt positiv ausfallen wird. Ungeachtet der Leistungsfähigkeit existierender Suchmaschinen zeigen Untersuchungen, dass erstens Suchmaschinen prinzipbedingt beileibe nicht alles finden können – große Teile des Internet und des World Wide Web sind für sie und damit für die Nutzerinnen und Nutzer von Suchmaschinen unsichtbar. Im nächsten Abschnitt wird daher zunächst die prinzipielle Funktionsweise von Suchmaschinen skizziert, da sich daraus bereits ergibt, aus welchen Gründen heraus bestimmte Teile des Netzes für Suchmaschinen unsichtbar sind und warum sich die in Suchmaschinen genutzten Algorithmen sowohl als deren Stärke als auch als deren Achillesferse erweisen.

Danach wird darauf einzugehen sein, dass die Ergebnislisten, die Suchmaschinen liefern, beileibe nicht die Ansprüche erfüllen können, die sowohl die Nutzerinnen und Nutzer erwarten bzw. die Suchmaschinenbetreiber versprechen. Die Ursachen hierfür

<sup>4</sup> Angaben jeweils in Prozent, siehe <<http://www.luna-park.de/home/internet-fakten/suchmaschinen-marktanteile.html>>, Werte aus 2009, zuletzt besucht am 24.08.2010.

sind vielfältig: Die technische Realisierung der Suchmaschinen, also die genutzten Algorithmen, beabsichtigte und unbeabsichtigte Folgen des Nutzerverhaltens, unterschiedlich begründete redaktionelle Eingriffe vonseiten der Suchmaschinenbetreiber sowie Wechselwirkungen zwischen diesen Faktoren beeinflussen die Ergebnisse von Suchmaschinen. Die Ergebnisse sind aufgrund dieser Einflüsse verzerrt – im Englischen wird hier von ‚search engine bias‘ gesprochen. All dies wird in den Abschnitten 3 und 4 angesprochen.

Den Schluss dieses Beitrags werden einige Handreichungen für Maßnahmen markieren, mit denen das Problem des search engine bias verkleinert werden kann. Grundsätzlich muss aber schon hier betont werden, dass die Hoffnung auf eine vollständige Abbildung der Internetinhalte mithilfe von Suchmaschinen eine Utopie ebenso wie eine Dystopie ist und bleiben wird. Suchmaschinen sind nicht die allwissenden Orakel, für die sie oft gehalten werden. Das ist Chance und Problem zugleich (vgl. [9]).

## 2. Funktionsweise von Suchmaschinen

Es gibt verschiedene Möglichkeiten, eine Suchmaschine für das Internet zu kreieren (vgl. [10]). Als sich Mitte der 1990er Jahre das World Wide Web zu etablieren begann, wurden Suchmaschinen mit Telefonbüchern oder Branchenverzeichnissen verglichen – tatsächlich gab es sogar einmal sogenannte Yellow Pages – also Gelbe Seiten – in gedruckter Form für das WWW. Auch die oft zu hörende Metapher der digitalen Bibliothek legt nahe, dass man die Inhalte des Internet oder des WWW mithilfe eines Katalogs zugänglich machen könnte. Das bedeutet, dass entsprechend der Inhalte einer Webseite diese verschlagwortet und in Kategorien eingeteilt wird und dann einen entsprechenden Katalogeintrag bekommt. Den so entstehenden Katalog könnten die Nutzerinnen und Nutzer in der Folge durchsuchen und auf diese Weise gewünschte Inhalte finden. Tatsächlich ging bspw. *Yahoo* in den 1990er Jahren diesen Weg und teilte Webseiten in einem von Menschen durchgeführten redaktionellen Prozess in Kategorien ein. Daher war *Yahoo* zu diesem Zeitpunkt auch keine Suchmaschine im engeren Sinne, denn diese werden üblicherweise dadurch charakterisiert, dass sie einen ‚Crawler‘ zum Durchsuchen des Internet nutzen (vgl. [11], 39ff. und [12]) – *Yahoo* war ein Webverzeichnis. Diese Art der Katalogisierung ist jedoch nur solange möglich, wie die Zahl der Webseiten halbwegs überschaubar ist; heute ist dies nicht mehr gegeben.

Stattdessen werden die bereits genannten ‚Crawler‘ bzw. ‚Webcrawler‘ genutzt. Dies sind letztlich Programme bzw. Algorithmen, die systematisch die Inhalte von Webseiten durchsuchen, Schlüsselwortlisten für die Indexierung der Webseiten generieren und die darin enthaltenen Schlüsselworteinträge mit den Adressen der Webseiten verbinden. Die Qualität einer Suchmaschine bzw. ihrer Suchergebnisse ist mit der Aktualität des Indexes gekoppelt (vgl. [13] und [14]); je öfter der Index aktualisiert wird, desto genauer können die Suchergebnisse sein. Als weitere Qualitätsindikatoren nennt Lewandowski ([15], 245) die Qualität der Treffer in den Ergebnislisten, die Recherchequalität, die sich bspw. durch die Möglichkeit komplexer Abfragemöglichkeiten äußert, sowie die Gestaltung der Benutzeroberfläche einer Suchmaschine. Wang, Xie und Goh [16] führen vergleichbare Kriterien an; außerdem nennen sie die schnelle Reaktion auf Suchanfragen. Elizabeth van Couvering [17] zeigt allerdings auf, dass die Sichtweise von Verantwortlichen bei Suchmaschinenbetreibern auf die Frage nach Qualität sehr stark markt- bzw. technologiegetrieben ist und sich nicht unbedingt an den Wünschen der Nutzerinnen und Nutzer orientiert.

Das bloße Zählen von Worten bringt aber auch Nachteile mit sich, da es nicht offensichtlich ist, wie man die Relevanz einer Seite in Bezug auf Suchausdrücke feststellen könnte. Zählt der Algorithmus bspw. die Worthäufigkeit auf einer Seite, so ist dies sehr manipulationsanfällig: Der Besitzer einer Webseite kann einfach bestimmte Schlüsselwörter sehr oft aufführen – möglicherweise für einen menschlichen Betrachter der Seite sogar unsichtbar – um so die Relevanz seiner Seite zu erhöhen. Dies ermöglicht gezielte Optimierung,<sup>5</sup> zuweilen könnte man auch von Manipulation und Missbrauch sprechen – weiter unten wird dazu im Zusammenhang mit dem ‚google bombing‘ bzw. ‚search engine bombing‘ noch mehr zu sagen sein.

Solche Eingriffsmöglichkeiten rufen nach Verbesserung. Die entscheidende Innovation, die *Googles* Suchmaschine nun mit sich brachte, ist, dass der dort genutzte ‚PageRank‘-Algorithmus (siehe bspw. [18], [19] und [20]) ein weiteres Kriterium zur Relevanzbestimmung nutzt: Bei seiner Suche im Internet zählt *Googles* Webcrawler die Zahl der Links, die auf eine Webseite verweisen. Je öfter also eine Webseite von anderen Seiten verlinkt wurde, desto relevanter scheint sie zu sein. Allerdings ist wichtig, dass die Verlinkungshäufigkeit nicht das einzige Kriterium für die Bestimmung der Relevanz einer Webseite und damit deren Position in der Ergebnisliste ist – es fließt eine Vielzahl von weiteren Faktoren ein, über die *Google* – für andere Anbieter von Suchmaschinen gilt dies ebenfalls – jedoch keine detaillierte Auskunft gibt.<sup>6</sup> Daraus folgt, dass alle Überlegungen über die Ursachen von Verzerrungen in den Ergebnissen von Suchmaschinen auf – im besten Fall – plausiblen Hypothesen beruhen, ohne dass diese direkt durch Inspektion der Suchalgorithmen überprüft werden könnten. Dies schränkt die Validität der folgenden Aussagen ohne Zweifel ein, doch sind die beschriebenen Verzerrungen real und die dafür angebotenen Erklärungen gut belegt.

### 3. Was Suchmaschinen (prinzipiell) nicht finden können

Eine Suchmaschine, die mit einem Webcrawler arbeitet, erstellt also einen Index, der Suchworte und Webseiten verbindet und zugleich die Relevanz einer bestimmten Webseite in Bezug auf ein bestimmtes Suchwort ermittelt und im Index abspeichert. Dieser Vorgang findet unabhängig von den Suchanfragen der Internetnutzerinnen und -nutzer statt. Die gigantischen Rechnerfarmen, die Suchmaschinenanbieter betreiben, sind also ständig damit beschäftigt, das Internet systematisch zu durchforsten und dabei den Index der entsprechenden Suchmaschine auf den neuesten Stand zu bringen. Natürlich muss dabei das Ziel angestrebt werden, dass der verwendete Webcrawler möglichst alle Inhalte des Netzes erfasst und in den Index aufnimmt. Es wäre allerdings ein großer Irrtum zu glauben, dass dies gelingt oder zumindest im Prinzip gelingen könnte. Denn tatsächlich stehen Webcrawlern zahlreiche Hindernisse entgegen, die eine vollständige Erfassung des Netzes und seiner Inhalte verhindern.

---

<sup>5</sup> ‚Search engine optimization‘ (SEO) ist hier das Stichwort, das zu einer Vielzahl von Fundstellen bei einer Suche führt (siehe bspw. [21] und [22]).

<sup>6</sup> Die Informationen, die bspw. Google preisgibt (siehe <<http://www.google.com/corporate/tech.html>>, zuletzt besucht am 17.12.2010), sind sehr allgemein gehalten.

### 3.1. Gezielter Ausschluss von Suchmaschinen

Eine dieser Hürden ist, dass gezielt verhindert werden kann, dass Seiten in den Index von Suchmaschinen aufgenommen werden. Dazu gibt es mehrere Techniken: Zum einen kann man im Wurzelverzeichnis einer Webpräsenz eine Datei ‚robots.txt‘ anlegen, die Anweisungen enthält, ob es Webcrawlern erlaubt sein soll, diese Webpräsenz zu durchsuchen. Es ist möglich, dies sehr differenziert zu gestalten, da die Verbote auf bestimmte Verzeichnisse, Dateien oder auch auf bestimmte Webcrawler beschränkt werden können.<sup>7</sup> Die Beachtung der Anweisungen in der Datei ‚robots.txt‘ durch Webcrawler wird allerdings nicht durch Gesetze, DIN- oder ISO-Normen oder beliebig andere verpflichtende Regeln vorgeschrieben; es gibt nicht einmal einen RFC<sup>8</sup> als Designrichtlinie für den Umgang mit dieser Datei. Die Befolgung der darin enthaltenen Anweisungen beruht auf einer stillschweigenden Übereinkunft, die aktuell verfügbare Suchmaschinen aber einzuhalten scheinen – zumindest tauchen entsprechend geschützte und damit ausgeschlossene Teile einer Webpräsenz nicht in den Ergebnislisten auf.

Ähnliches gilt für die zweite Technik, die Nutzung eines sogenannten Meta-Tags oder Meta-Elements im HTML-Code jeder einzelnen Webseite (vgl. [23], 100 ff.). Hiermit kann man erneut angeben, ob Webcrawlern erlaubt sein soll, eine bestimmte Webseite in den Index aufzunehmen; außerdem kann bestimmt werden, wie mit den Links auf einer Webseite verfahren werden soll.<sup>9</sup>

In sehr vielen Fällen werden Inhalte, die durch die Nutzung solcher Techniken von der Indexierung durch Suchmaschinen ausgeschlossen werden, tatsächlich keine oder kaum Relevanz für Suchmaschinen besitzen, weil die entsprechenden Verzeichnisse bspw. nur temporäre Dateien oder Ähnliches enthalten. Doch grundsätzlich muss klar sein, dass durch Nutzung solcher Ausschlussmechanismen Suchmaschinen nicht überall Zugang im Netz finden und somit nicht alle Inhalte des Internet indexieren können.<sup>10</sup>

### 3.2. Inselbildung und das unsichtbare Netz

Mit der Nutzung solcher Ausschlussverfahren kann ein weiterer Effekt einhergehen, den man als ‚Inselbildung‘ bezeichnen könnte. Denn die Verwendung der Datei ‚robots.txt‘ und von Meta-Tags bzw. Meta-Elementen erlaubt es, Webseiten komplett der Sichtbarkeit für Suchmaschinen zu entziehen. Diese Webseiten können durchaus mit anderen Seiten verlinkt und damit untereinander verbunden sein, doch da Meta-Tags bzw. Meta-Elemente die Verfolgung solcher Links durch Suchmaschinen verbieten können, werden sie von diesen nicht indexiert. In diesen Fällen ‚sehen‘ Suchmaschinen also weniger als die Benutzerinnen und Benutzer des Internet. Eine noch extremere Form der Inselbildung liegt vor, wenn Webseiten vollständig isoliert sind, also keine Links auf sie verweisen, sondern Nutzerinnen und Nutzer diese nur direkt durch Eingabe der entsprechenden URL aufrufen können. Liegen solche Seiten in Verzeichnissen, die

<sup>7</sup> Im WWW finden sich zahlreiche Beschreibungen, Erklärungen und Gebrauchsanleitungen zum genauen Aufbau und zur Verwendung der Datei ‚robots.txt‘; einen guten Einstieg bietet bspw. <<http://de.selfhtml.org/diverses/robots.htm>>, zuletzt besucht am 27.01.2011.

<sup>8</sup> RFC = Request for Comments, siehe <<http://tools.ietf.org/rfc/index>>, zuletzt besucht am 11.12.2010.

<sup>9</sup> Auch hier bieten wieder die SELFHTML-Seiten einen guten Einstieg in die Funktionsweise, siehe <<http://de.selfhtml.org/html/kopfdaten/meta.htm>>, zuletzt besucht am 27.01.2011.

<sup>10</sup> Theo Röhle ([23], 102) bemerkt dazu: „Allerdings bleibt die Entscheidung darüber, welche Meta-Tags ausgewertet werden, komplett Google überlassen. Von Seiten der Inhalteanbieter bestehen keine Verhandlungsmöglichkeiten, um die Auswertung weiterer Meta-Tags von Google einzufordern.“

durch Einträge in der Datei ‚robots.txt‘ für Suchmaschinen gesperrt sind, bleiben sie sowohl für Suchmaschinen unsichtbar als auch für alle Nutzerinnen und Nutzer, die über keine Kenntnis der richtigen URL verfügen.

Inhalte, die von Suchmaschinen nicht gefunden und damit nicht indexiert werden, werden im Englischen als ‚deep web‘, ‚deep net‘, ‚invisible web‘, ‚dark net‘, ‚dark web‘ oder ‚hidden web‘ bezeichnet; die für Suchmaschinen sichtbaren Teile wiederum werden im Englischen mit dem Ausdruck ‚surface web‘ bezeichnet. Das deep web besteht nicht nur aus den gerade schon genannten Inseln; nach Schätzungen ist es um mehrere Größenordnungen umfangreicher als das surface web (siehe [24] sowie kritisch dazu [25]; vgl. [23], 88 ff.; eine Fallstudie für den Tourismusbereich bietet [26]). Die Ursachen für die Entstehung des deep web sind zahlreich und unterschiedlich:

- Inseln aufgrund fehlender Verlinkung,
- Inseln aufgrund der Nutzung von Meta-Elementen, der Datei ‚robots.txt‘ oder anderer Zugangshürden wie Captchas,
- passwortgeschützte Seiten,
- dynamisch erzeugte Webseiten z. B. aufgrund von Nutzereingaben in Suchmasken,
- Inhalte des Internet, die von Suchmaschinen nicht durchsucht werden (können), weil sie in einem der Suchmaschine nicht bekannten Format gespeichert sind.

Diese Liste ist vermutlich nicht vollständig (für eine systematische Unterteilung siehe [27]); aufgrund immer neuer Technologien, die genutzt werden, um Inhalte an die Nutzerinnen und Nutzer des Internet auszuliefern, macht es wenig Sinn, hier Vollständigkeit anzustreben. Entscheidend ist, dass eine Vielzahl von Faktoren die Entstehung und das Anwachsen des deep web fördert. Aus Nutzersicht bedeutet die Existenz des deep web, dass große Teile des WWW nicht durch Suchmaschinen erschlossen werden können; Ergebnislisten von Suchmaschinen sind also grundsätzlich defizitär bzw. verzerrt, da sie die Inhalte des deep web nicht enthalten, keine Angaben darüber machen können, welche Relevanz diese Inhalte für eine gegebene Suchanfrage hätten und auch nicht beziffern können, wie umfangreich die nicht gefundenen Inhalte wären. In solchen Fällen, die vermutlich die Regel und nicht die Ausnahme darstellen, müssen Nutzerinnen und Nutzer des Internet auf andere Suchstrategien zurückgreifen – dazu wird am Ende des Textes noch etwas zu sagen sein.

Natürlich gibt es Bestrebungen, das deep web über Standardsuchmaschinen zugänglich zu machen, da dies einen großen Wettbewerbsvorteil für eine Suchmaschine bedeutete, denn die Qualität der Ergebnislisten könnte so erheblich verbessert werden. In der entsprechenden Literatur finden sich hierzu zahlreiche Ansätze, dieses Ziel zu erreichen (bspw. [28]). Allerdings werden einige der Faktoren, die oben genannt wurden, aus grundsätzlichen Erwägungen nicht durch technische Verfahren zu überwinden sein, weil hierbei auch rechtliche Aspekte zu bedenken sind. Wenn bspw. Inhalte des Internet durch Passwörter, Meta-Elemente oder Ähnliches geschützt werden, so bedeutet dies ein klares Signal an Suchmaschinen: In solchen Fällen wollen jene Personen, Unternehmen oder Institutionen, die diese Webseiten betreiben, eben nicht, dass die entsprechenden Inhalte allgemein zugänglich sind und im Index von Suchmaschinen auftauchen. Das heißt, dass bestimmte Teile des deep web für Suchmaschinen prinzipiell stets unzugänglich bleiben sollen. Es ist daher auch fraglich, ob dann von einem search engine bias gesprochen werden sollte. Denn in jenen Fällen, in denen gezieltes Benutzerverhalten Suchmaschinen vom Indexieren von Inhalten abhält, sollen die entsprechenden Inhalte ausdrücklich nicht gefunden werden – sie sind damit nicht öffentlich zugänglich.



Man könnte nun argumentieren, dass dieses Nichterfassen im positiven Sinne ein Qualitätskriterium einer Suchmaschine sei, da diese den Wünschen der Inhaltsanbieter folgt.

### 3.3. Was Suchmaschinen nicht finden wollen

Sieht man von der freiwilligen Selbstbeschränkung von Suchmaschinen durch die Beachtung von Meta-Tags bzw. Meta-Elementen und der Datei ‚robots.txt‘ ab, war bisher immer nur die Rede davon, was Suchmaschinen nicht finden *können*; es gibt aber auch Inhalte im Netz, die Suchmaschinen nicht indexieren und damit auch nicht finden *wollen*. Denn selbst wenn die Vollständigkeit der Suchergebnisse ein wichtiges Qualitätskriterium für Suchmaschinen ist, gibt es noch weitere Eigenschaften, anhand derer man Suchmaschinen beurteilen kann: Übersichtlichkeit der Suchergebnisse, keine irrelevanten Ergebnisse und keine Dopplungen in den Ergebnislisten usw. – diese Kriterien um der Vollständigkeit willen zu ignorieren könnte nachteilige Wirkungen auf den Erfolg einer Suchmaschine haben. Außerdem ist es fraglich, ob eine vollständige Abdeckung des WWW – unter der eher zweifelhaften Annahme der technischen Umsetzbarkeit – ökonomisch sinnvoll wäre. Denn dies bedeutete, dass Suchmaschinenbetreiber zu noch größeren Investitionen in ihre Infrastruktur genötigt wären als heute – große Unternehmen wie *Google*, die durch Werbung enorme Einnahmen generieren, mögen solche Kosten stemmen können, kleinere Unternehmen wären damit aber vermutlich überfordert.

Die vollständige Abdeckung brächte aber auch für die Nutzerinnen und Nutzer von Suchmaschinen nicht nur Vorteile, sondern in vielen Fällen Nachteile. So darf bspw. die Online-Enzyklopädie *Wikipedia* frei kopiert werden – was auch sehr oft passiert. Würden Suchmaschinen dies nicht berücksichtigen, wären deren Ergebnislisten voll Doppelungen bzw. Dubletten, da alle Kopien eines Artikels der *Wikipedia* aufgelistet würden – aus Nutzersicht nicht unbedingt ein Mehrwert. Ähnliches gilt für alle Inhalte des WWW, die oft kopiert werden, bspw. Pressemeldungen und Produktankündigungen von Unternehmen, Blogeinträge oder Erfahrungsberichte über den Einsatz eines Computers oder anderer Konsumgüter.

Aus Sicht der Suchmaschinenbetreiber macht es daher Sinn, solche Inhalte zumindest in den Ergebnislisten nicht aufzuführen, da eine Anzeige zur Unübersichtlichkeit, aber nicht zu einer Qualitätssteigerung der Ergebnisse in Hinsicht auf Vollständigkeit und Relevanz beitrüge. Ob Suchmaschinenbetreiber solche nicht angezeigten Inhalte in irgendeiner Form trotzdem verarbeiten, ist jedoch eine andere Frage. Es wäre denkbar, dass die Webcrawler von Suchmaschinen Auswertungen generieren, die für statistische Zwecke genutzt werden und bspw. zur Verbesserung der Platzierung von Werbung nützlich sein könnten.

### 3.4. Vergleichbarkeit von Suchmaschinen

Das bisher Gesagte zeigt bereits, dass die Gestaltung einer Suchmaschine ein komplexes Unterfangen ist; umso mehr stellt sich die Frage nach einem einfachen Maß, mit dem verschiedene Suchmaschinen verglichen werden könnten. Es liegt nahe, dieses Maß in der Größe des Indexes einer Suchmaschine zu sehen – je größer der Index, desto leistungsfähiger die Suchmaschine, da ein größerer Index eine bessere Abdeckung des WWW anzudeuten scheint. Tatsächlich finden sich zuweilen Angaben über die Indexgröße von Suchmaschinen, doch ist dies kein taugliches Maß zur Bestimmung der Qualität. Denn eine Suchmaschine, die bspw. Dubletten, Spam und andere Inhalte ohne Unterschied indexieren würde, hätte zwar einen umfangreichen Index, aber lieferte nicht notwendigerweise gute Ergebnisse.

#### 4. Verzerrung der Ergebnislisten

Wirft man einen Blick auf die Ergebnislisten verschiedener Suchmaschinen, fällt sofort auf, dass zwar Übereinstimmungen vorliegen, sich aber auch deutliche Unterschiede zeigen. Dies bezieht sich sowohl auf die Inhalte der Ergebnisliste als auch auf die Reihenfolge der einzelnen Listeneinträge. Für die Nutzerinnen und Nutzer von Suchmaschinen stellt sich daher die Frage, welche Suchmaschine denn nun die richtigen Ergebnisse liefert. Einer Antwort auf diese Frage nähert man sich bereits dadurch, dass man überlegt, was in diesem Zusammenhang denn ‚richtig‘ bedeutet.

##### 4.1. Ein kleines Experiment

Die folgenden Ergebnislisten (Tabellen 2 und 3) der Suchmaschinen *Ask*, *Bing*, *Google* und *Yahoo* wurden alle am 09.12.2010 zwischen 13:00 und 14:00 Uhr in kurzer zeitlicher Abfolge produziert. Die Ergebnislisten wurden soweit bearbeitet, dass sie hier gut dargestellt werden können; inhaltlich wurden Änderungen nur insoweit vorgenommen, dass Angaben gelöscht wurden, die für das diskutierte Thema nicht relevant sind. Jede der folgenden vier Listen gibt die erste Seite der Fundstellen wieder, die eine Suche nach „search engine bias“ (die Anführungszeichen sind wichtig) erzeugte. Eine Reproduktion dieser Ergebnislisten zu einem späteren Zeitpunkt wird nicht möglich sein, da diese sich mit der Zeit verändern; außerdem wird *Yahoo* in Zukunft *Bing*, die Suchmaschine von *Microsoft*, nutzen;<sup>11</sup> *Ask* wiederum wird den Betrieb seiner Suchmaschine in absehbarer Zeit ganz einstellen.<sup>12</sup>

Ein erster Vergleich der Suchergebnisse von *Ask* und *Bing* zeigt bereits erhebliche Unterschiede: Die inhaltliche Überschneidung der beiden Listen ist klein; auch die Positionierung der Einträge, die in beiden Listen enthalten sind, differiert.

Der direkte Vergleich der Suchergebnisse von *Google* und *Yahoo* zeigt ebenfalls erhebliche Unterschiede in Bezug auf die gefundenen Einträge und deren Positionierung. Stünden alle vier Ergebnislisten nebeneinander, würde der paarweise Vergleich der Ergebnisse diese Unterschiede für alle Suchmaschinen deutlich sichtbar werden lassen.<sup>13</sup> Genau diese Unterschiede werden nun als Verzerrung bzw. search engine bias bezeichnet. Als empirische Tatsache ist diese Verzerrung vermutlich den meisten Nutzerinnen und Nutzern des Internet und von Suchmaschinen bekannt; die dahinter liegenden Ursachen sind jedoch nicht nur rein technische im Sinne bspw. der Leistungsfähigkeit der jeweiligen Suchmaschineninfrastruktur. Es gibt bessere und schlechtere Suchmaschinen, aber es gibt keine Suchmaschine, die den Stein der Weisen gefunden hätte.

<sup>11</sup> Siehe dazu <<http://www.heise.de/newsticker/meldung/Websuche-bei-Yahoo-USA-auf-Microsoft-Bing-gestellt-1065614.html>>, zuletzt besucht am 17.12.2010.

<sup>12</sup> Siehe dazu <<http://www.heise.de/newsticker/meldung/Aus-fuer-die-Suchmaschine-Ask-com-1133997.html>>, zuletzt besucht am 17.12.2010.

<sup>13</sup> Schon dies legt nahe, dass man mehrere Suchmaschinen nutzen sollte, um möglichst vollständige Suchen durchführen zu können. Allerdings setzt das voraus, dass verschiedene Suchmaschinen Ergebnisse mit hinreichender Nichtüberdeckung liefern. Rather, Lone und Shah [29] zeigen in einer Studie, dass diese Forderung dann leicht zu erfüllen ist, wenn man allgemeine Suchmaschinen wie *Bing* oder *Google* mit spezialisierten Suchmaschinen ergänzt; allerdings schränken die Autoren ihren Befund insoweit ein, als dass sie nur einen bestimmten Themenbereich untersucht haben. Spink et al. [30] kommen zu einem Ergebnis, das den hier dokumentierten Selbsttest bestätigt: Auch die Ergebnislisten allgemeiner Suchmaschinen überlappen sich ihrer Studie zufolge nur sehr wenig, so dass der Einsatz von mehreren Suchmaschinen bzw. von Metasuchmaschinen sehr sinnvoll erscheint.

**Tabelle 2:** Ergebnisliste von Ask und Bing

	Ask	Bing
1.	<b>Are Search Engines Biased? - Search Engine Watch (SEW)</b> 11 Mar 2002 ... Can you trust search engines to deliver only fair and objective results? Maybe not, say the authors of a study seeking to measure how bias ...	<b>Search Engine Bias</b> Microsoft PowerPoint – Venn
2.	<b>View or Download - Search Engine Bias: Evidence and Possible Causes</b> studied search engine bias by testing whether one search engine retrieves results that are significantly different from those of a control group of search ...	<b>Trust and the Internet: Search Engine...</b> SEO by the Sea. Internet Marketing and Search Engine Optimization (SEO) Services, Consulting, and Research
3.	<b>Google: Advertising and search engine bias   ZDNet</b> 19 Dec 2006 ... Sergey Brin and Larry Page on how an advertising-based business model can negatively effect search engine quality.	<b>Search engine bias</b> Search engines are not biased toward popular and highly linked websites, researchers report, in this week's PNAS. The internet is ...
4.	<b>Measuring search engine bias</b> Analyzing search engine bias. In Proceeding of the 1st international conference on computing and information technologies (ICCIT), Montclair, NJ (pp. 3-8). ...	<b>SEARCH ENGINE BIAS DEMISE OF SEARCH ENGINE UTOPIANISM</b> S EARCH E NGINE B IAS AND THE D EMISE OF S EARCH E NGINE U TOPIANISM 189 In the past few years, search engines have emerged as a major force in our information economy, helping searchers ...
5.	<b>Trust and the Internet: Search Engine Bias</b> 15 May 2006 ... But, during a workshop on Models of Trust for the Web, there's a paper being presented that takes a close look at search engine bias, ...	<b>Search engine bias - François ...</b> Slashdot is wondering how objective is Microsoft's search: There have been a number of stories on Microsoft trying to do a 'Netscape' on Google ... what would a world in which ...
6.	<b>Search engine bias</b> 8 Aug 2006 ... Search engines are not biased toward popular and highly linked websites, researchers report, in this week's PNAS. The internet is ...	<b>Search Engine Bias General Search Engine...</b> search engine bias ...
7.	<b>View PDF - SEARCH ENGINE BIAS AND THE DEMISE OF SEARCH ENGINE ...</b> Instead, this Essay argues that search engine bias is the ... Search engine bias sounds scary, but this Essay explains why such bias is both necessary and ...	<b>SSRN-Search Engine Bias and the Demise of...</b> SSRN-Search Engine Bias and the Demise of Search Engine Utopianism by Eric Goldman
8.	<b>InformIT: Search Engine Bias and the Demise of Search Engine ...</b> 22 Sep 2006 ... Some commentators believe that search engine bias is a defect requiring legislative correction. But Eric Goldman argues that search engine ...	<b>Search engine</b> Search engine bias. Although search engines are programmed to rank websites based on their popularity and relevancy, empirical studies indicate various political, economic, and ...
9.	<b>MediaPost Publications Deconstructing Search Engine Bias 02/27/2008</b> 27 Feb 2008 ... For a thorough discussion of the major issues surrounding algorithmic bias, I highly recommend reading Eric Goldman's "Search Engine Bias ...	<b>InformIT: Search Engine Bias and the...</b> Due to automated operations, people often mistakenly assume that search engines display search results neutrally and without bias. Like any other media company, however, search ...
10.	<b>Search-Engine Bias Project</b> Search-Engine Bias Project. Search engines are a big part of our everyday life. Most of us rely on search engines to discover and access contents from the ...	<b>Search Engine Bias and SEO</b> There is some controversy brewing between SEO Specialists that there is no bias within the search engines themselves. This I am here to tell you that this is true. Bias runs wild ...

**Tabelle 3:** Ergebnisliste von *Google* und *Yahoo*

	<i>Google</i>	<i>Yahoo</i>
1.	<b>Are Search Engines Biased? - Search Engine Watch (SEW)</b> 11 Mar 2002 ... Can you trust search engines to deliver only fair and objective results? Maybe not, say the authors of a study seeking to measure how bias ...	<b>Position Paper: A Study of Web Search Engine Bias and its Assessment</b> Search engine bias has been seriously noticed in recent years. ... engines is termed search engine bias. ... only provides a partial view of search engine bias. ...
2.	<b>Trust and the Internet: Search Engine Bias</b> 15 May 2006 ... But, during a workshop on Models of Trust for the Web, there's a paper being presented that takes a close look at search engine bias, ...	<b>Are Search Engines Biased? - Search Engine Watch (SEW)</b> Can you trust search engines to deliver only fair and objective results? ... What is a reasonable conversion expectation from click to form fi...
3.	<b>Search engine bias</b> 8 Aug 2006 ... Search engines are not biased toward popular and highly linked websites, researchers report, in this week's PNAS. The internet is ...	<b>Web search engine - Wikipedia, the free encyclopedia</b> A web search engine is designed to search for information on the World Wide Web and FTP servers. The search results are generally presented in a list of results and are often called hits. The information may consist of web...
4.	<b>View PDF - SEARCH ENGINE BIAS AND THE DEMISE OF SEARCH ENGINE ...</b> von E GOLDMAN - Zitiert durch: 47 Instead, this Essay argues that search engine bias is the ... Search engine bias sounds scary, but this Essay explains why such bias is both necessary and ...	<b>Measuring search engine bias</b> Measuring search engine bias. Abbe Mowshowitz , Akira Kawaguchi. 1 ... Analyzing search engine bias. In Proceeding of the 1st international conference on computing ...
5.	<b>MediaPost Publications Deconstructing Search Engine Bias 02/27/2008</b> 27 Feb 2008 ... For a thorough discussion of the major issues surrounding algorithmic bias, I highly recommend reading Eric Goldman's "Search Engine Bias ...	<b>Search Engine Bias and the Demise of Search Engine Utopianism</b> SSRN-Search Engine Bias and the Demise of Search Engine Utopianism by Eric Goldman ... Some commentators believe that search engine bias is a defect requiring ...
6.	<b>Search-Engine Bias Project</b> Search-Engine Bias Project. Search engines are a big part of our everyday life. Most of us rely on search engines to discover and access contents from the ...	<b>InformIT: Search Engine Bias and the Demise of Search Engine Utopianism ...</b> Search Engine Bias Is Necessary and Desirable ... Some commentators believe that search engine bias is a defect requiring legislative correction. ...
7.	<b>View or Download - Search Engine Bias: Evidence and Possible Causes</b> von L Vaughan - Zitiert durch: 135 studied search engine bias by testing whether one search engine retrieves results that are significantly different from those of a control group of search ...	<b>Michael Zimmer.org " Search Engine Bias</b> China, Google, Search Engine Bias, Search Engines " ... Google News, Search Engine Bias, Values in Design " Transparency in GoogleNews ...
8.	<b>InformIT: Search Engine Bias and the Demise of Search Engine ...</b> 22 Sep 2006 ... Some commentators believe that search engine bias is a defect requiring legislative correction. But Eric Goldman argues that search engine ...	<b>Search Engine Bias: Evidence and Possible Causes</b> School of Computing and Information Technology, University of ... studied search engine bias by testing whether one search engine retrieves results that are ...
9.	<b>Google: Advertising and search engine bias   ZDNet</b> 19 Dec 2006 ... Sergey Brin and Larry Page on how an advertising-based business model can negatively effect search engine quality.	<b>Search Engine Bias, Chapter 2</b> As discussed in Chapter 1, Section 1.2, search engine bias can be both ... develop a perspective which stresses the conflict latent in the discussion of search engine ...
10.	<b>Search Engine Bias, Chapter 1</b> scholarly attention to the idea of search engine bias, as well as ... may play a role in search engine bias (cf. Vaughan & Thelwall, 2004) and the location ...	<b>Search Engine Bias</b> Advanet Pty Ltd, SEO Australia Consultant, Solutions for website visibility and accessibility, SEO ... See if you can find other examples of search engine bias. ...

#### 4.2. Definitionen

Bevor jedoch auf die zahlreichen Ursachen für die Unterschiede der einzelnen Ergebnislisten eingegangen werden kann, muss zunächst einmal geklärt werden, was unter einer Verzerrung bzw. einem search engine bias verstanden werden soll. Ein erster Schritt dazu ist, eine knappe Definition für Suchmaschinen selbst zu geben ([31], 141): „Information retrieval systems are designed to furnish information in response to queries; they organize collections of items such as business records, medical files, scientific articles, etc. by indexing collections, providing a language for asking questions, and a retrieval algorithm for extracting responses to queries. Search engines are information retrieval systems whose collection consists of pages of the World Wide Web.”

Über die Funktionsweise von aktuellen Suchmaschinen ist weiter oben schon kurz etwas gesagt worden. Unklar ist aber noch, was es heißt, dass die Ergebnisliste einer Suchmaschine verzerrt ist ([31], 142): „Bias in retrieval systems is evidenced by undue inclusion or exclusion of certain items among those retrieved in response to queries; or it is revealed in giving undue prominence to some items at the expense of others. [...] An unbiased system should retrieve a selection that neither exaggerates nor downplays any particular items in the database in response to a set of queries.“ Die Über- bzw. Unterrepräsentation von Webseiten findet sich auch in anderen Definitionen für die Verzerrung von Suchergebnissen, doch können sich die Bezugsgrößen, anhand derer die Verzerrung gemessen wird, unterscheiden ([35], 2): „[...], “bias” is conceptually defined as a non-proportional coverage of a particular group of Web sites, e.g. sites from a particular country. Under representation of a group of sites by a search engine is considered a bias against these sites by the engine while over representation is viewed as a bias in favour of the sites in question.” Diese Definition von Vaughan und Thelwall deutet bereits an, dass die auf Webseiten verwendete Sprache zu Verzerrungen beitragen kann. Man kann Verzerrung allerdings auch ex negativo bestimmen ([15], 246, kursiv im Original): „Idealerweise würde ein Suchergebnis aus allen relevanten Treffern, die im Datenbestand enthalten sind, bestehen. Es würde also einerseits *nur* relevante Treffer enthalten, andererseits *alle* relevanten Treffer [...] zurückgeben.“ Ist diese Forderung nicht erfüllt, liegt eine Verzerrung vor.

Es ist nun aber kaum möglich, bspw. per Hand eine Suche im Internet durchzuführen, um eine unverzerrte Ergebnisliste zu erzeugen, die dann mit der Ergebnisliste einer Suchmaschine verglichen wird, um zu bestimmen, wie groß die Verzerrung jener Suchmaschine ist – diese Vorgehensweise ist allein schon aus praktischen Erwägungen heraus nur in Spezialfällen gangbar. Daher muss ein anderer Weg beschritten werden, der durch die oben gezeigten Ergebnislisten von vier existierenden Suchmaschinen bereits angedeutet wurde: Es wird ein Vergleich zwischen den Ergebnissen verschiedener Suchmaschinen gezogen ([31], 142): „We propose to measure bias by comparing the performance of one retrieval system to that of a group of such systems, all of which function in the same information universe.“ Allerdings ist dies nicht der einzige mögliche Weg, Verzerrungen zu messen; im nächsten Abschnitt werden andere Verfahren kurz beschrieben.

Dieser Vergleichsaspekt ist nicht nur für die Bestimmung des Verzerrungsgrades von Ergebnislisten wichtig, sondern deutet auf ein viel weitergehendes Problem hin: Wenn es in Zukunft nur noch eine Suchmaschine geben sollte, also bspw. *Google* alle anderen Suchmaschinen vom Markt verdrängt haben sollte, könnte man nicht mehr sinnvoll von Verzerrungen der Ergebnislisten sprechen, da ein Vergleichsmaßstab fehlte – zumindest müsste ein ganz anderes Maß dafür herangezogen werden. Allein um die

Qualität von Suchmaschinen beurteilen zu können, sollte es daher immer eine Alternative geben, auf die man ausweichen kann. Denn Suchmaschinen haben eine gesellschaftliche Dimension, wie Mowshowitz und Kawaguchi ([31], 142) betonen, obwohl es ihnen zunächst um im weitesten Sinne technische Aspekte von Suchmaschinen geht: „Individuals search the Web for information pertaining to work, leisure, entertainment, current events, personal finance, politics, etc. As dependence on the Web grows, so grows the need for reliable and – especially in the realm of citizenship – unbiased information. Decision makers in private organizations and public agencies also have a need for unbiased information, or at least knowledge of the bias of their sources.“<sup>14</sup> Ähnliche Bemerkungen finden sich auch bei anderen Autoren; meist wird in diesem Zusammenhang auf die Gatekeeperfunktion von Suchmaschinen hingewiesen (bspw. [8]).

## 5. Ursachen für die Ergebnisverzerrung von Suchmaschinen

Die Bedeutung von Suchmaschinen zur Erschließung der Inhalte des Internet ist, folgt man der obigen Stellungnahme, also kaum zu überschätzen. Ungeachtet dieser Bedeutung lässt sich aber schon jetzt festhalten, dass es grundsätzlich keine unverzerrten Suchergebnisse gibt und geben kann; insbesondere dann nicht, wenn man die prinzipbedingten Lücken der Suche, wie sie oben beschrieben wurden, in Rechnung stellt. Wenn es aber keine unverzerrten Suchergebnisse geben kann, wäre es zumindest hilfreich, wenn es möglich wäre, den Grad der Verzerrung qualitativ feststellen zu können. Sollte auch dies nicht möglich sein, wäre die Kenntnis der Ursachen der Verzerrungen wichtig, da dies die Entwicklung von Gegenstrategien aufseiten der Nutzerinnen und Nutzer erlauben würde.

Abschätzungen der Verzerrungen in den Ergebnissen von Suchmaschinen können hier nicht vorgenommen werden; hierfür sei auf die entsprechende Literatur verwiesen (siehe bspw. [31], [32], [33], [34], [35] und [36]). Daher kann an dieser Stelle selbst eine qualitative Angabe des Verzerrungsgrades nicht geboten werden. Aber die Ursachen für die Verzerrungen sind recht gut untersucht; dies erlaubt es, zum Schluss dieses Textes einige Strategien zu skizzieren, um hiermit umzugehen. Methoden zur Untersuchung von Verzerrungen sind aufwändig: Brown, Dubois und Shepard [33] bspw. führten im Bereich der Paläontologie über mehrere Jahre eine umfangreiche Quellenrecherche zu bestimmten Amphibien aus dem Perm durch, um dann ihre eigenen Ergebnisse mit jenen von Suchmaschinen zu vergleichen; Cho und Roy [34] ebenso wie Vaughan und Thelwall [32] kopierten definierte Bereiche des WWW, suchten in den Kopien nach Ausdrücken und verglichen die so generierten Ergebnisse mit den entsprechenden Ergebnislisten von Suchmaschinen; Mowshowitz und Kawaguchi ([31], [35]) kombinierten aus den Ergebnissen mehrerer Suchmaschinen eine ideale Ergebnisliste und maßen dann die Abweichungen jeder einzelnen Suchmaschine gegenüber dem Ideal. Uyar [36] ging es vor allem darum zu überprüfen, ob die Ergebnislisten von Suchmaschinen tatsächlich so viele Einträge beinhalten, wie als geschätzte Anzahl an Dokumenten am Beginn der Ergebnisliste angegeben wird. Dazu wurden Programme geschrieben, die automatisch Suchausdrücke (in diesem Fall Zahlen) generierten, Suchmaschinen damit aufriefen und dann die Ergebnislisten durchliefen, um

<sup>14</sup> Beispiele eindeutiger und nachgewiesener Verzerrungen sind allerdings schwer zu finden. Benjamin Edelman und Benjamin Lockwood (vgl. <<http://www.benedelman.org/searchbias/>>, zuletzt besucht am 27.01.2011) behaupten, solche Fälle nachweisen zu können, doch sind die Suchausdrücke, um die es dabei geht, nicht unbedingt von gesellschaftlicher Bedeutung in einem engeren Sinne.

die geschätzte Dokumentenzahl mit der tatsächlichen zu vergleichen. Jede dieser Methoden hat Vor- und Nachteile und deren Ergebnisse können selbst verzerrt sein: Es gibt bspw. keine Garantie dafür, dass eine manuelle Quellenrecherche ein vollständiges Ergebnis liefert; die ausgewählten Bereiche des WWW könnten nicht repräsentativ für das WWW sein; die Zusammenfassung der Ergebnisse verschiedener Suchmaschinen wird kaum den tatsächlichen Inhalt des WWW wiedergeben (siehe dazu die Anmerkungen zum deep web); die Suche nach Zahlen erscheint artifiziell und hat wenig mit realen Suchanfragen zu tun. Daher sind Aussagen über das Ausmaß der Verzerrung in den Ergebnissen von Suchmaschinen mit Vorsicht zu genießen.

Die Ursachen für Verzerrungen können – obwohl es auch Überschneidungen gibt – in drei Kategorien eingeteilt werden: Solche, die in der Implementierung (in einem weiten Sinne) der jeweiligen Suchmaschine liegen, solche, die durch das Verhalten der Anbieter von Inhalten im Internet zustande kommen und schließlich Ursachen, die darin liegen, wie Suchmaschinen genutzt werden. Natürlich gibt es auch Kombinationen dieser Ursachen. Die folgenden Anmerkungen streben keine Vollständigkeit an, sondern sollen dafür sensibilisieren, dass die Ergebnislisten von Suchmaschinen nicht von einem gutwilligen allwissenden Orakel erstellt werden, sondern Produkt einer Vielzahl interdependenter Faktoren sind, die zu einem erheblichen Anteil die divergierenden Interessen unterschiedlicher Akteure im Bereich der Suchmaschinennutzung reflektieren.

### 5.1. Implementierung

Dass die Implementierung der Suchalgorithmen erheblichen Einfluss auf die Leistungsfähigkeit einer Suchmaschine hat, wurde bereits angesprochen. Daher sollen nun Aspekte der Implementierung angesprochen werden, die weniger offensichtlich sind, aber trotzdem beträchtlichen Einfluss auf Suchergebnisse haben können.

Suchmaschinen für das Internet müssen eine gigantische Informationsmenge bewältigen – jede Optimierung, die zur Komplexitätsreduktion bei der Reaktion auf Suchanfragen beiträgt, ist daher sowohl aus technischer als auch ökonomischer Sicht willkommen; zudem könnten womöglich die Suchergebnisse verbessert werden. Ein solcher Mechanismus ist nun das im Englischen mit ‚stemming‘ bezeichnete Verfahren, nicht jede Wortform zu indexieren, sondern möglichst nur den Wortstamm – da dies für alle Information Retrieval-Systeme nützlich sein kann, begannen entsprechende Forschungen schon recht früh (vgl. [37]). Stemming könnte zur Umfangsreduzierung des Suchindexes beitragen, vor allem aber erlaubt es Suchmaschinen, auf natürlichsprachige Eingaben besser zu reagieren: Die eingegebenen Suchworte werden auf ihren Wortstamm reduziert und dann wird nach diesem Wortstamm gesucht; im besten Fall sind damit alle Wortformen, also bspw. Deklinationen und Konjugationen, durch diese Suche abgedeckt. Soll eine Suchmaschine für verschiedene Sprachen mithilfe von stemming Suchergebnisse liefern, müssen die spezifischen sprachlichen Eigenheiten beachtet werden – hier muss also erhebliches linguistisches Wissen in Algorithmen gegossen werden. Die Implementierung solcher Algorithmen entscheidet nun mithin darüber, ob stemming die Suchergebnisse verbessert oder verzerrt. Am Beispiel von *Google* zeigt Ahmet Uyar [38], dass die hier verwendeten Algorithmen erheblichen Einfluss darauf haben, welche Inhalte gefunden werden und auf welcher Position in der Ergebnisliste sie erscheinen. Es gibt also einen Trade-off zwischen der Unterstützung natürlichsprachiger Sucheingaben und der Verzerrung der Ergebnislisten. Es ist denkbar, dass in heutigen Suchmaschinen weitere Mechanismen dieser Art realisiert sind, die ähnliche

Wirkungen zeitigen. Interessant an der Untersuchung Uyars ([38], 500 f.) ist, dass zumindest *Google* sich bei der Nutzung von stemming gegen eine Reduzierung der Indexgröße entschieden hat; tatsächlich scheint *Google* nicht nur den Wortstamm zu indexieren, sondern jede einzelne Wortform. Dies vergrößert zwar den Index, erlaubt aber vermutlich schnellere Antwortzeiten bei Suchanfragen.

Eine völlig andere Ursache für die Verzerrung der Ergebnislisten sind rechtliche Vorgaben. Nicht nur in China werden Suchmaschinen reguliert und darüber die Anzeige von bestimmten Inhalten verhindert; dies geschieht auch in Deutschland und vielen anderen Ländern. So haben sich Suchmaschinenanbieter in Deutschland unter dem Dach des Vereins *Freiwillige Selbstkontrolle Multimedia-Diensteanbieter* (FSM) einem Verhaltenskodex unterworfen, der unter anderem beinhaltet, dass Webseiten, die von der *Bundesprüfstelle für jugendgefährdende Medien* (BPJM) indiziert wurden, nicht in den Ergebnislisten der betreffenden Suchmaschinen erscheinen.<sup>15</sup> Es ist unerheblich, ob dies als gerechtfertigte Maßnahme bspw. gegen jugendgefährdende oder extremistische Inhalte oder als Zensur angesehen wird; entscheidend ist, dass die Ergebnislisten von Suchmaschinen gezielt beeinflusst werden, um bestimmte Inhalte nicht anzuzeigen.

Für diese Einflussnahme muss eine entsprechende technische Infrastruktur geschaffen werden, die natürlich auch zu anderen Zwecken genutzt werden kann. Die Betreiber von Suchmaschinen könnten über die Gestaltung ihrer Algorithmen darauf Einfluss nehmen, dass Angebote von Konkurrenzunternehmen nicht oder auf einer unattraktiven Position angezeigt werden.<sup>16</sup> Derzeit untersucht die EU-Kommission mögliche Kartellrechtsverstöße durch *Google*.<sup>17</sup> Auch wenn es hierbei zunächst um Verdachtsmomente geht, muss klar sein, dass Suchmaschinenbetreiber erhebliche Möglichkeiten besitzen, gezielt auf Ergebnislisten Einfluss zu nehmen.

## 5.2. Anbieterverhalten

Informationen zu suchen ist teuer; in der Ökonomie spricht man in diesem Zusammenhang von ‚Suchkosten‘ bzw. ‚Informationskosten‘. Suchen bedeutet Ressourcenverbrauch, vor allem Zeitaufwand. Dies führt dazu, dass viele Nutzerinnen und Nutzer von Suchmaschinen nur die erste Seite einer Ergebnisliste zur Kenntnis nehmen (vgl. [39]); die übrigen Fundstellen werden also weitgehend ignoriert.<sup>18</sup> Gerade für Unternehmen, die ja vom Verkauf eines Produktes und/oder einer Dienstleistung leben, ist es daher überlebenswichtig, in den Suchergebnissen auf der ersten Seite zu stehen, wenn jemand Ausdrücke zur Suche verwendet, die mit jenen Produkten und/oder Dienstleistungen in Verbindung stehen. Es ist daher aus Unternehmenssicht naheliegend, die eigenen Seiten dahingehend zu optimieren, dass diese bei möglichst vielen infrage kommenden Suchausdrücken sehr weit oben in der Ergebnisliste einer Suchmaschine aufgeführt werden. Dies wird, wie weiter oben bereits angesprochen, als Suchmaschinenoptimierung bzw. *search engine optimization* bezeichnet. Der Buchmarkt ebenso wie das Internet selbst bevorratet eine Fülle von Ratschlägen, wie Webseitenanbieter vorgehen müssen, um ihre Seiten auf Sichtbarkeit und Relevanz hin zu optimieren.

<sup>15</sup> Siehe <[https://www.fsm.de/de/Selbstkontrolle\\_Suchmaschinen](https://www.fsm.de/de/Selbstkontrolle_Suchmaschinen)>, zuletzt besucht am 18.12.2010.

<sup>16</sup> Siehe dazu bspw. <<http://www.benedelman.org/hardcoding/>> und <<http://www.benedelman.org/searchbias/>>, beide zuletzt besucht am 27.01.2011.

<sup>17</sup> Siehe <<http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/1624&format=PDF&aged=0&language=DE&guiLanguage=en>>, zuletzt besucht am 18.12.2010.

<sup>18</sup> Andere Unterlassungen wie die Nichtbenutzung von Suchoperatoren wie ‚and‘, ‚or‘ und ‚not‘ haben jedoch kaum Einfluss auf die Ergebnisse (vgl. [40]).



Es ist sicherlich legitim, dass sich Webseitenanbieter Gedanken darüber machen, welche Suchausdrücke auf ihre Produkte und/oder Dienstleistungen hinweisen, um diese dann gezielt auf den eigenen Seiten aufzuführen. Davon ist aber abzugrenzen, dass es Webseiten gibt, die eine Fülle von Wörtern enthalten, nach denen häufig gesucht wird, ohne dass dies mit dem sichtbaren Inhalt der Webseite oder den angebotenen Produkten und/oder Dienstleistungen etwas zu tun hätte – es geht hier nur darum, Nutzerinnen und Nutzer auf diese Seiten zu locken. Search engine optimization kann also gezielt missbraucht werden – eine klare Grenze zwischen legitimer Optimierung und missbräuchlicher Gestaltung ist allerdings nicht zu ziehen. Da moderne Suchmaschinen die Relevanz einer Seite auch nach anderen Kriterien ermitteln und dabei der PageRank-Algorithmus eine wichtige Rolle spielt, ist eine weitere Möglichkeit, Ergebnisse einer Suchmaschine zu manipulieren, sogenannte Linkfarmen anzulegen: Dies sind Webseiten, die eine Fülle von Links auf andere Seiten enthalten. Dies soll den PageRank-Faktor und damit die Relevanz der verlinkten Seiten erhöhen – in diesem Fall ist der Missbrauch offensichtlich. Die Betreiber von Suchmaschinen versuchen allerdings, solche Maßnahmen zu erkennen und ihre Algorithmen entsprechend zu gestalten. Neuere Mechanismen, die zur Suchmaschinenoptimierung von Webseiten genutzt werden, sind bspw. ‚scraper sites‘, ‚link baiting‘ oder ‚Linkkauf‘. Link baiting hat dabei große Ähnlichkeit mit dem ‚search engine bombing‘ (s.u.) und wird häufig im viralen Marketing verwendet; so kann bspw. ein Unternehmen durch Produktankündigungen, die auf News-Seiten oder in Blogs reproduziert werden, viele Links auf die eigenen Webseiten erzeugen.

Doch Webseiten sollen zuweilen nicht nur besonders sichtbar sein; oftmals werden sie so gestaltet, dass sie gerade nicht in den Ergebnislisten von Suchmaschinen auftauchen. Die so verborgenen Inhalte gehören dem deep web an (s.o.); sie zu erschließen soll in der Regel nur über die Suchmöglichkeiten des entsprechenden Contentproviders möglich sein – bspw. im Fall von Onlineausgaben von Zeitungen mit Premium- und/oder kostenpflichtigen Angeboten. Es geht hierbei meist um die Wahrung kommerzieller Interessen in einem doppelten Sinne: Erstens sollen die Einnahmen aus kostenpflichtigen Angeboten geschützt werden, zweitens sollen Einnahmen aus Werbung auf den eigenen Seiten generiert werden.

### 5.3. Nutzerverhalten<sup>19</sup>

Die Ergebnislisten von Suchmaschinen geben nicht unverzerrt wieder, welche Inhalte im Internet zu finden sind – das zeigen die Tabellen 2 und 3 sehr deutlich. Die Tabellen weisen aber noch etwas auf, das eigentlich offensichtlich ist und gerade deshalb unsichtbar bleibt: Die Inhalte, auf die die Ergebnislisten verlinken, sind alle in englischer Sprache verfasst. Wie könnte es auch anders sein, mag man fragen, da doch ein englischer Suchausdruck verwendet wurde? Es mag trivial erscheinen, dass eine englischsprachige Suche nur englischsprachige Ergebnisse liefert, doch muss klar sein, dass die Beschränkung einer Suche auf Suchbegriffe einer Sprache erheblich zur Verzerrung der Ergebnisse beitragen kann (vgl. bspw. [41]).<sup>20</sup> Denn selbst wenn es keine anderen Ursachen für Verzerrungen gäbe, bedeutete eine entsprechende Beschränkung auf eine

<sup>19</sup> Hier ist bewusstes Verhalten gemeint; es gibt andere Effekte aufseiten der Nutzerinnen und Nutzer, die zu einer Fehlwahrnehmung der Ergebnisse einer Suchmaschine beitragen können (vgl. bspw. [42]).

<sup>20</sup> Allerdings spricht die Möglichkeit des search engine bombing (s.u.) dafür, dass bspw. englischsprachige Seiten mit deutschen Suchausdrücken gefunden werden würden, sofern die Verlinkung in deutscher Sprache gehalten ist (vgl. auch [43]).

Sprache, dass in der Regel ein großer Teil der im Internet vorhandenen Informationen nicht gefunden würde.<sup>21</sup>

Gerade in der universitären Lehre begründen viele Studierende die Nutzung zweifelhafter Quellen oft damit, dass sie nichts anderes gefunden hätten. Sicher mag es noch andere Ursachen geben, aber wenn man mit diesen Studierenden ihre Recherche-strategie durchspielt, zeigt sich meist, dass sie bspw. nur in deutscher Sprache und mit einer Standardsuchmaschine wie *Google* gesucht haben. In vielen Fällen muss dies scheitern, da die entsprechende wissenschaftliche Debatte in Englisch stattfindet. Außerdem wird die Beschränkung auf eine Standardsuchmaschine nicht der Tatsache gerecht, dass viele Inhalte des Netzes im deep web liegen und so nicht gefunden werden können, sondern nur durch die Nutzung von bereichsspezifischen Suchmaschinen.

#### 5.4. Kombinationen

Die in den drei vorhergehenden Abschnitten aufgeführten Ursachen für die Verzerrung der Ergebnisse von Suchmaschinen konnten mehr oder minder deutlich entweder der Implementierung, dem Verhalten von Webseitenanbietern oder dem Nutzerverhalten zugeordnet werden. Doch ist es kaum überraschend, dass es auch Kombinationen dieser Ursachen gibt.

Eine solche Kombination ist die schon mehrfach angesprochene Inselbildung, die mithin dazu führt, dass Inhalte des Internet zum deep web gehören. Inseln können aber nur entstehen, weil Suchmaschinen Maßnahmen von Webseitenanbietern gegen die Indexierung ihrer Seiten respektieren. Es gibt eine Vielzahl von Möglichkeiten, die Anbieter von Webseiten nutzen können, um ihre Inhalte der Sichtbarkeit durch Suchmaschinen wie *Google* zu entziehen. Wie schon bemerkt gibt es Bemühungen, einige dieser Möglichkeiten zu umgehen – aber selbst wenn entsprechende Techniken verfügbar sein sollten, stünde deren Anwendung immer noch unter rechtlichen Vorbehalten, da Inhaltsanbieter diese Maßnahmen ja ergreifen, um ihr geistiges Eigentum zu schützen (dazu bspw. [44]). Ähnlich wie im Fall der Filterung von bestimmten Seiten aus den Ergebnislisten von Suchmaschinen spielt Recht also auch in diesem Fall eine erhebliche Rolle in Bezug auf die Verzerrung von Suchergebnissen.

Zum Schluss soll eine Verzerrungsursache angesprochen werden, die im Englischen als ‚search engine bombing‘ oder – um der Marktmacht von *Google* gerecht zu werden – ‚Google bombing‘ bezeichnet wird. Dabei wird ausgenutzt, dass Suchmaschinen Webseiten nicht nur mit Inhalten der Webseite selbst indexieren, sondern auch mit den Inhalten von anderen Seiten, auf denen sich Links zur zu indexierenden Webseite befinden: Wenn jemand auf einer Webseite über Suchmaschinenverzerrung auf eine Webseite über search engine bias verlinkt, wird die Suche nach ‚Suchmaschinenverzerrung‘ auch die Seite zu search engine bias finden. Clifford Tatum beschreibt nun eine politische Nutzung des search engine bombing [45]: „Left-wing oriented bloggers linked the words “miserable failure” to Bush’s bio page – at <http://www.whitehouse.gov/president/gwbbio.html> – on the whitehouse.gov Web site and eventually increased its Google P Ranking to the number one slot resulting from a search of the words “miserable failure.” Not satisfied to let their leader be defaced in cyberspace, right-wing bloggers began a campaign to displace Bush as the number one ranked “miserable

<sup>21</sup> Sucht man mit *Google* bspw. nach ‚Suchmaschinenverzerrung‘ als wörtliche Übersetzung von ‚search engine bias‘, werden nur zwei Einträge gefunden; sucht man hingegen mit ‚Suchmaschinen Verzerrung‘, ist zwar die Ergebnisliste sehr lang, aber oftmals haben die Einträge nichts mit dem Gesuchten zu tun.

failure” by collective linking, or Google bombing, the Web sites of Michael Moore, Jimmy Carter, and Hillary Clinton.“ Die Nutzerinnen und Nutzer des Internet können also durch koordiniertes Handeln die Eigenheiten der Implementierung von Suchmaschinen für politische Kampagnen ausnutzen. Denn sie können in Blogs oder auf Webseiten bspw. die Seiten einer Person oder Institution verlinken und dabei den Inhalt ihrer eigenen Webseite so gestalten, dass sie ihre Meinung über jene Person oder Institution zum Ausdruck bringen. James Grimmelman ([46], 942 f., kursiv im Original) bringt die Bedeutung dieser Möglichkeit auf den Punkt: „This is a significant new form of politicking. Land a bomb like this and you can convince the world that Google agrees with your position. A successful Googlebomb doesn’t just *reflect* the consensus of web users; it can help *construct* that consensus.“ Damit schließt sich der Kreis zu dem Zitat von Mowshowitz und Kawaguchi ([31], 142): „As dependence on the Web grows, so grows the need for reliable and – especially in the realm of citizenship – unbiased information. Decision makers in private organizations and public agencies also have a need for unbiased information, or at least knowledge of the bias of their sources.“ Der vorliegende Text sollte zumindest die letzte Forderung erfüllen – nämlich Hinweise auf die Ursachen für die Verzerrungen in den Ergebnislisten von Suchmaschinen zu liefern.

## 6. Dr. Seltsam, oder wie ich lernte, die Suchmaschine zu lieben

Auf die Nutzung von Suchmaschinen kann heute niemand mehr verzichten – sie sind im Guten wie im Schlechten Gatekeeper zu den Inhalten des Internet. Lawrence Hinman beschreibt das Problem dieser Funktion als Gatekeeper mit folgenden Worten ([47], 21): „[Search engines] are like windows onto the web – and, like windows, tend to be largely unnoticed because our gaze focuses on what is visible through them. With windows, however, it is easy to detect when they are cloudy or distorted. With search engines, however, it is much more difficult to tell when they are providing distorted or incomplete pictures.“ Doch einige Möglichkeiten gibt es, die Verzerrungen zu erkennen; manche wurden bereits angedeutet: Solange es Alternativen gibt, sollte man diese nutzen. Dies gilt nicht nur für die Konkurrenz zwischen *Google* und *Bing*, den einzigen relevanten Wettbewerbern auf dem Suchmaschinenmarkt, wenn man einmal von einigen länderspezifischen Ausnahmen wie *Baidu* (siehe Tabelle 1) absieht, sondern insbesondere für die Nutzung bereichsspezifischer Suchmaschinen bspw. von Bibliotheken, Verlagen und anderen Informationsanbietern. Gerade im Bildungsbereich ist es wichtig, Schülern und Schülerinnen sowie Studierenden nahezubringen, dass sich eine Recherche nicht auf die Nutzung einer Suchmaschine oder der *Wikipedia* beschränken darf. Zumindest die teilweise Erschließung des deep web ist durch die Verwendung bereichsspezifischer Suchmaschinen möglich und für ernsthaftes wissenschaftliches oder journalistisches Arbeiten auch unverzichtbar.<sup>22</sup>

Darüber hinaus ist bei jeder Suche wichtig, sich darüber bewusst zu sein, dass Suchergebnisse vielfältigen Einflüssen ausgesetzt sind. Zuweilen lassen sich diese

---

<sup>22</sup> Eine Fallstudie, die dies sehr gut demonstriert, findet sich in [50], siehe außerdem den Beitrag zur journalistischen Recherche im vorliegenden Sammelband. Als Beispiele wissenschaftlicher Suchmaschinen können ‚Base‘ und ‚Scirus‘ (<<http://www.base-search.net>> und <<http://www.scirus.com>>, beide zuletzt besucht am 28.01.2011) genannt werden. Margaret Marklands [51] Studie deutet ebenfalls an, dass die ausschließliche Nutzung von Suchmaschinen wie *Google* zu nicht optimalen Ergebnissen im Bereich wissenschaftlicher Recherche führt.

durch Umwege in der Recherche umgehen, doch es bleibt festzuhalten, dass es unverzerrte Suchergebnisse prinzipiell nicht geben kann. Schaut man sich die oben dokumentierten Suchergebnisse für „search engine bias“ an, werden insbesondere die Unterschiede in Hinblick auf die Positionierung von Einträgen deutlich – es lohnt sich daher, nicht nur die erste Seite der Ergebnisliste anzuschauen.

Die Möglichkeiten der massiven Einflussnahme auf Suchergebnisse, wie sie sich bspw. im search engine bombing manifestieren, weisen auf ein allgemeines Problem im Umgang mit den Inhalten des Netzes hin: Viele Nutzerinnen und Nutzer tendieren dazu, nur Inhalte wahrzunehmen, die ihre jeweils eigene Meinung bestätigen – sie vermeiden damit kognitive Dissonanz (vgl. bspw. [48] und [49]). Suchmaschinen können diesen Trend verstärken.<sup>23</sup> Ist man sich dieses Effekts bewusst, lässt er sich mit wenig Aufwand vermeiden. Darüber hinaus ist wichtig, die eigene Sprache nicht als Nabel der Welt zu sehen: Gerade im wissenschaftlichen Kontext werden die meisten wichtigen Debatten zu großen Teilen in englischer Sprache geführt. All diese Umgehungsstrategien können helfen, ändern jedoch nichts an der grundsätzlichen Problematik. Wirklichen bedenklich wäre die Situation aber, wenn es auf dem Suchmaschinenmarkt keinen Wettbewerb mehr gäbe. Dann wären supranationale Regulierungsmaßnahmen notwendig. Doch ob diese überhaupt erreicht werden könnten, steht auf einem anderen Blatt. Vor allem aber wären staatliche Eingriffe möglicherweise eine Kur, die schlimmer ist als die Krankheit, denn es gibt wenig Grund zur Annahme, dass staatliche Eingriffe nicht selbst zu massiven Verzerrungen der Ergebnislisten von Suchmaschinen führen würden.

## Literatur

- [1] T. Röhle, Desperately seeking the consumer: Personalized search engines and the commercial exploitation of user data, *First Monday* 12 (9), (frei zugänglich unter <[http://firstmonday.org/issues/issue12\\_9/rohle/index.html](http://firstmonday.org/issues/issue12_9/rohle/index.html)>, zuletzt besucht am 19.12.2010), 2007.
- [2] O. Tene, What Google Knows: Privacy and Internet Search Engines. In: *Utah Law Review* 4, 2008, 1433-1492.
- [3] M. Zimmer, The Externalities of Search 2.0: The Emerging Privacy Threats when the Drive for the Perfect Search Engine meets Web 2.0, *First Monday* 13 (3), (frei zugänglich unter <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2136/1944>>, zuletzt besucht am 19.12.2010), 2008.
- [4] C. Shapiro & H. R. Varian, *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, Boston/Massachusetts, 1998.
- [5] PEW, Search Engine Users Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve. *PEW Internet & American Life Project* (frei zugänglich unter <[http://www.pewinternet.org/~media/Files/Reports/2005/PIP\\_Searchengine\\_users.pdf.pdf](http://www.pewinternet.org/~media/Files/Reports/2005/PIP_Searchengine_users.pdf.pdf)>, zuletzt besucht am 19.12.2010), 2005.
- [6] M. Machill & M. Beiler, Die Bedeutung des Internets für die journalistische Recherche, *Media Perspektiven* 10, 2008, 516-531 (frei zugänglich unter <[http://www.media-perspektiven.de/uploads/tx\\_mppublications/10-2008\\_Machill\\_Beiler.pdf](http://www.media-perspektiven.de/uploads/tx_mppublications/10-2008_Machill_Beiler.pdf)>, zuletzt besucht am 19.12.2010).
- [7] E. Hargittai, The Social, Political, Economic, and Cultural Dimensions of Search Engines: An Introduction, *Journal of Computer-Mediated Communication* 12 (3), 2007, 769-777.

---

<sup>23</sup> Eine neue Suchmaschine mit dem Namen ‚Blekko‘ (<[www.blekko.com](http://www.blekko.com)>, zuletzt besucht am 28.01.2011) kann ebenfalls zur Verminderung kognitiver Dissonanz genutzt werden, denn man kann von vornherein die Art der Suchergebnisse festlegen. Das auf den Webseiten dieser Suchmaschine selbst genutzte Beispiel ist die Suche nach ‚climate warming /conservative‘: Mit dem Wort nach dem Schrägstrich gibt man an, dass man nur Ergebnisse sehen will, die eine konservative Position (im Sinne der US-amerikanischen politischen Szene) einnehmen. Dies erlaubt eine sehr individuelle Suche, doch gleichzeitig kann die Perspektive der Suchenden verengt werden.

- [8] M. Machill, C. Neuberger, W. Schweiger & W. Wirth, Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen. In: M. Machill & C. Welp (Hrsg.): *Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen*, Bertelsmann Stiftung, Gütersloh, S. 13-490, 2003.
- [9] E. Goldman, Search engine bias and the demise of search engine utopianism, *Yale Journal of Law and Technology* **8**, 2005, 188-200.
- [10] J. Griesbaum, B. Bekavac & M. Rittberger Typologie der Suchdienste im Internet. In: D. Lewandowski, (Hrsg.): *Handbuch Internet-Suchmaschinen. Nutzerorientierung in Wissenschaft und Praxis*. Akademische Verlagsanstalt AKA, Heidelberg, 2009, S. 18-52.
- [11] J. Battelle, *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. Portfolio, New York, 2005.
- [12] K. Patzwaldt, Suchmaschinenlandschaften. In: K. Lehmann & M. Schetsche (Hrsg.): *Die Google-Gesellschaft: Vom digitalen Wandel des Wissens*. Transcript, Bielefeld, 2005 S. 75-82.
- [13] D. Lewandowski A three-year study on the freshness of web search engine databases, *Journal of Information Science* **34** (6), 2008, 817-831.
- [14] D. Lewandowski, H. Wahlig & G. Meyer-Bautor, The freshness of web search engine databases, *Journal of Information Science* **32** (2), 2006, 131-148.
- [15] D. Lewandowski, Mit welchen Kennzahlen lässt sich die Qualität von Suchmaschinen messen? In: M. Machill & M. Beiler, (Hrsg.): *Die Macht der Suchmaschinen*. Halem, Köln, 2007, S. 243-258.
- [16] H. Wang, M. Xie, & T. N. Goh, Service quality of Internet search engines, *Journal of Information Science* **25** (6), 1999, 499-507.
- [17] E. van Couvering, Is Relevance Relevant? Market, Science, and War: Discourses of Search Engine Quality, *Journal of Computer-Mediated Communication* **12** (3), 2007, 866-887.
- [18] S. Brin & L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems* **30** (1), 1998, 107-117.
- [19] L. Pretto, A Theoretical Analysis of Google's PageRank, *String Processing and Information Retrieval, Lecture Notes in Computer Science* **2476**. Springer, Berlin Heidelberg, 2002, 125-136.
- [20] R. Wills, Google's pagerank, *The Mathematical Intelligencer* **28**, 2006, 6-11.
- [21] R. A. Malaga, Worst practices in search engine optimization, *Communications of the ACM* **51** (12), 2008, 147-150.
- [22] B. Xing & Z. Lin, The impact of search engine optimization on online advertising market, *Proceedings of the 8<sup>th</sup> international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet*. Frederickton, New Brunswick, Canada: ACM, S. 519-529, 2006.
- [23] T. Röhle, *Der Google-Komplex. Über Macht im Zeitalter des Internets*, Transcript, Bielefeld, 2010.
- [24] M. K. Bergman, White Paper: The Deep Web: Surfacing Hidden Value, *Journal of Electronic Publishing* **7** (1), 2001 (frei zugänglich unter <<http://hdl.handle.net/2027/spo.3336451.0007.104>>, zuletzt besucht am 19.12.2010).
- [25] D. Lewandowski & P. Mayr, Exploring the academic invisible web, *Library Hi Tech* **24** (4), 2006, 529-539.
- [26] Z. Xiang, K. Wober, & D. R. Fesenmaier, Representation of the online tourism domain in search engines, *Journal of Travel Research* **47** (2), 2008, 137-150.
- [27] C. Sherman & G. Price The Invisible Web. *Uncovering Information Sources Search Engines Can't See*. Information Today, Medford, 2001.
- [28] J. Madhavan, D. Ko, K. Lucja, V. Ganapathy, A. Rasmussen & A. Halevy, Google's Deep Web crawl, *Proceedings of the VLDB Endowment* **1** (2), 2008, 1241-1252.
- [29] R. A. Rather, F. A. Lone & G. J. Shah, Overlap in Web Search Results: A Study of Five Search Engines, *Library Philosophy and Practice* **2008**, 2008, 1-6.
- [30] A. Spink, B. J. Jansen, C. Blakely & S. Koshman, A study of results overlap and uniqueness among major Web search engines, *Information Processing & Management* **42** (5), 2006, 1379-1391.
- [31] A. Mowshowitz & A. Kawaguchi, Assessing bias in search engines, *Information Processing & Management* **38** (1), 2002, 141-156.
- [32] L. Vaughan & M. Thelwall, Search engine coverage bias: evidence and possible causes, *Information Processing & Management* **40** (4), 2004, 693-707.
- [33] L. E. Brown, A. Dubois & D. B. Shepard, Inefficiency and Bias of Search Engines in Retrieving References Containing Scientific Names of Fossil Amphibians, *Bulletin of Science, Technology & Society* **28** (4), 2008, 279-288.
- [34] J. Cho & S. Roy, Impact of search engines on page popularity, *Proceedings of the 13<sup>th</sup> international conference on World Wide Web*, ACM, New York, 2004, 20-29.
- [35] A. Mowshowitz & A. Kawaguchi, Bias on the Web, *Communications of the ACM* **45** (9), 2002, 56-60.
- [36] A. Uyar, Investigation of the accuracy of search engine hit counts, *Journal of Information Science* **35** (4), 2009, 469-480.

- [37] J. B. Lovins, Development of a Stemming Algorithm. In: *Mechanical Translation and Computational Linguistics* **11** (1), 1968, 22-31.
- [38] A. Uyar, Google stemming mechanisms, *Journal of Information Science* **35** (5), 2009, 499-514.
- [39] W. Wirth, T. Böcking, V. Karnowski & T. von Pape, Heuristic and Systematic Use of Search Engines, *Journal of Computer-Mediated Communication* **12** (3), 2007, 778-800.
- [40] C. M. Eastman & B. J. Jansen, Coverage, relevance, and ranking: The impact of query operators on Web search engine results, *ACM Transactions on Information Systems* **21** (4), 2003, 383-411.
- [41] J. Bar-Ilan & T. Gutman, How do search engines respond to some non-English queries?, *Journal of Information Science* **31** (1), 2005, 13-28.
- [42] Y. Kammerer & P. Gerjets, How the interface design influences users' spontaneous trustworthiness evaluations of web search results: comparing a list and a grid interface, *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. Austin, Texas, ACM, 2010, 299-306.
- [43] L. Vaughan & Y. Zhang, Equal representation by search engines? A comparison of websites across countries and domains, *Journal of Computer-Mediated Communication* **12** (3), 2007, 888-909.
- [44] B. Allgrove, The search engine's dilemma: implied licence to crawl and cache? In: *Journal of Intellectual Property Law & Practice* **2** (7), 2007, 437-438.
- [45] C. Tatum, Deconstructing Google bombs: A breach of symbolic power or just a goofy prank?, *First Monday* **10** (10), 2005, (frei zugänglich unter <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1287/1207>>, zuletzt besucht am 19.12.2010).
- [46] J. Grimmelmann, The Google Dilemma, *New York Law School Law Review* **53**, 2009, 939-950.
- [47] L. M. Hinman, Esse est indicato in Google: Ethical and Political Issues in Search Engines, *IRIE – International Review of Information Ethics* **3** (6), 2005, 19-25 (frei zugänglich unter <[http://www.i-r-i-e.net/inhalt/003/003\\_hinman.pdf](http://www.i-r-i-e.net/inhalt/003/003_hinman.pdf)>, zuletzt besucht am 19.12.2010).
- [48] K. Weber, R. Drücke, O. Langewitz & M. Nagenborg, Konvergente Medien – Integration oder Fragmentierung von Öffentlichkeit?, *Merz Medien + Erziehung, Zeitschrift für Medienpädagogik* **53** (6), 2009, 83-92.
- [49] K. Weber & R. Drücke, Konvergente Medien, Fragmentierung der Öffentlichkeit und medienethische Anmerkungen, *Zeitschrift für Kommunikationsökologie und Medienethik* **12** (1), 2010, 68-72.
- [50] J. Pearson & K. Soothill, Using an Old Search Engine: the Value of the Times Index, *Sociology* **37** (4), 2003, 781-790.
- [51] M. Markland, Does the student's love of the search engine mean that high quality online academic resources are being missed?, *Performance Measurement and Metrics* **6** (1), 2005, 19-31.