

# Suchmaschinen – Gatekeeper im Internet

Welche Rolle spielen Suchmaschinen für die Kommunikationsprozesse im Internet und welche Probleme sind damit verbunden? Um diese Fragen wird es im nachfolgenden Beitrag gehen. Es wird argumentiert, dass Suchmaschinen die Funktion von Gatekeepern haben, die darüber entscheiden, welche Informationen den Usern zugänglich sind und welche nicht. Aus diesem Grund ist die Qualität der Retrievals nicht nur aus der Perspektive von Suchmaschinenbetreibern, Content-Anbietern und Rezipienten von Bedeutung. Die kommunikative Funktion von Suchmaschinen, ihre Selektions- und Präsentationsroutinen sowie die Qualität der Rechercheergebnisse sollten auch aus gesellschaftlicher Perspektive diskutiert werden, denn bei genauerer Betrachtung zeigt sich, dass Suchmaschinen durchaus keine «neutralen Informationsvermittler» sind (vgl. Loosen 1999: 44; Kuhlen 1999: 78, 247).

## Von der Nutzerperspektive zu den Vermittlungsinstanzen: Suchmaschinen

Die Diskussion über die *gesellschaftlichen* Folgen der Internetkommunikation wird von der These der «*Digitalen Spaltung*» beherrscht. Im Kern geht es bei dieser These um die unterschiedlichen Chancen und Möglichkeiten, auf *Informationen* zuzugreifen, was auf die ungleiche Verteilung der Zugangswege zum Internet zurückgeführt wird. Die dazu vorliegenden empirischen Befunde zeigen, dass sowohl die Zugangsoptionen als auch der Umfang der Internetnutzung zwischen verschiedenen Bevölkerungssegmenten deutlich variieren (vgl. Kubicek & Welling 2000; Eimeren, Gerhard & Frees 2001; Vowe & Emmer 2001). Und nicht nur hinsichtlich der *generellen* Nutzung unterscheiden sich die soziodemographischen Gruppen, auch in Bezug auf die *Inhalte* der Nutzung gibt es Unterschiede: Von den Internetnutzern mit Hauptschulabschluss sucht beispielweise nur die Hälfte regelmäßig (mindestens einmal pro Woche) gezielt nach Informationen, bei den Befragten mit Studium sind es hingegen gut zwei Drittel (ARD-/ZDF-Online-Studie 2001).

Die Forschung, die sich mit der gesellschaftlichen Bedeutung und den Konsequenzen der Internetkommunikation zu *Informationszwecken* beschäftigt, wird bislang durch Analysen dominiert, die die Situation und die Entwicklung aus der Nutzerperspektive untersuchen. Im vorliegenden Beitrag soll diese Perspektive ergänzt werden: Neben den Anbietern von Informationen werden vor allem die Vermittlungsinstanzen im internetbasierten Informationsprozess im Zentrum der Betrachtung stehen, denn mit dem Zugang zum Internet sind die vielfältigen Informationsmöglichkeiten zwar theoretisch verfügbar aber für den praktischen Gebrauch bedarf es weiterer Hilfsmittel. Dies sind vor allem die Suchmaschinen. Für den gezielten Zugriff auf Informationen sind sie unverzichtbar und deswegen für den Kommunikationsprozess im Internet von strategischer Bedeutung. *Wer* darauf Einfluss hat, *welche* Informationen leicht, schwer oder gar nicht gefunden werden können und *wie* dieser Entscheidungsprozess verläuft, ist von *gesellschaftlicher* Relevanz. Suchmaschinen sind die Gatekeeper des Internets.<sup>1</sup> Ihre Such- und Auswahl-

routinen entscheiden darüber, welche Informationen den Rezipienten tatsächlich zugänglich sind. Von daher ist es wichtig, die Funktionsweisen und die Auswahlkriterien von Suchmaschinen einer kritischen Analyse zu unterziehen. Dabei eröffnen sich verschiedene Anknüpfungspunkte an theoretische Ansätze und Debatten: Nicht nur die Gatekeeperforschung ist hier zu nennen (vgl. Shoemaker 1991). Man stößt auch auf Phänomene, die dem Verhältnis zwischen Öffentlichkeitsarbeit und Journalismus nicht unähnlich sind (vgl. dazu Baerns 1987; Weischenberg 1995; Donsbach 1997). Auch die spannungsreiche Beziehung zwischen redaktionellem Angebot und Werbung spielt in diesem Zusammenhang eine Rolle.

## Die Akteure: Internetnutzer, Content-Anbieter, Suchmaschinenbetreiber

«Das Wettrüsten zwischen Sitebetreibern und Suchmaschinen wird vielleicht nie ein Ende finden». Mit diesem Satz beginnt ein Beitrag über die Kriterien, mit denen Suchmaschinen das Ranking für die gefundenen Webseiten durchführen (drweb.de 2001). Wie kommt der Autor des Beitrags dazu, das Verhältnis zwischen Suchmaschinenbetreibern und Content-Anbietern als «Wettrüsten» zu bezeichnen? Welche Interessen sind es, die zu einem solchen «Wettrüsten» führen? Um welche «Waffen» geht es bei diesem «Wettrüsten»?

Um das Konfliktpotential richtig einzuschätzen, muss man sich bewusst werden, welche Bedeutung die Suchmaschinen für die Kommunikation im Internet haben. Glaubt man der Selbsteinschätzung der *Internetnutzer*, dann ist die «zielgerichtete Suche nach Informationen» die am zweithäufigsten genutzte Anwendungsmöglichkeit des Internets, gleich nach dem «Empfangen und Versenden von E-Mails» (vgl. Eimeren, Gerhard & Frees 2001: 387). Eine *zielgerichtete* Suche kann über die direkte Eingabe einer bekannten WWW-Adresse erfolgen oder aber über den Einsatz von Suchmaschinen. Grundsätzlich werden beide Wege von den Usern intensiv genutzt. Wenn jedoch keine konkrete Adresse zum gesuchten Thema bekannt ist, dann ist der Rückgriff auf Suchmaschinen fast unvermeidlich (vgl. Loosen 1999: 42). Ohne ihre Hilfe wäre eine sinnvolle

Nutzung des WWW kaum möglich. Von daher ist es auch nicht erstaunlich, dass Suchmaschinen zu den meistgenutzten Angeboten im WWW gehören (vgl. V-business: Nutzung der Suchmaschinen).<sup>2</sup>

Suchmaschinen haben eine kaum zu überschätzende strategische Bedeutung für das Auffinden von Web-Angeboten. Insbesondere für die kommerziellen *Anbieter von Content* ist es von existenzieller Bedeutung, dass ihre Web-Seiten bei entsprechenden Suchanfragen von den Suchmaschinen auf den vorderen Rankingplätzen gelistet werden. Deswegen ist es nicht verwunderlich, dass es mittlerweile zahlreiche Internetfirmen gibt, die sich darauf spezialisiert haben, Web-Angebote dahingehend zu optimieren, dass sie von den verschiedenen Suchmaschinen prominent gerankt werden (vgl. Loosen 1999: 45). Angeboten werden von solchen Unternehmen u.a. die Ermittlung von zielgruppenspezifischen Suchwortstatistiken, Web-Seiten-Optimierungen, die Anmeldung bei geeigneten Suchmaschinen sowie kontinuierliche Erfolgskontrollen (vgl. als Beispiel: GAB+GESSINGER 2001). Neben den *Anbietern von Content* gibt es jedoch noch zwei weitere Akteure, deren Interessen man berücksichtigen muss, wenn man verstehen will, warum die Kommunikationsleistungen von Suchmaschinen umstritten und konfliktreich sind. Dabei handelt es sich zum einen um die *User* und zum anderen um die *Betreiber der Suchmaschinen*.

Welche Interessen haben diese Akteure? Fast alle Suchmaschinen werden von privatwirtschaftlich organisierten Unternehmen betrieben. Von daher sind die *Betreiber von Suchmaschinen* darauf angewiesen, mit ihren Angeboten Geld zu verdienen. Hierfür gibt es verschiedene Möglichkeiten: Sie können beispielsweise auf der Web-Seite der Suchmaschine *Bannerwerbung* verkaufen oder direkt eine Allianz mit einem Content-Anbieter eingehen. Des Weiteren kann für die Nutzung der Suchmaschine von den Usern eine *Nutzungsgebühr* erhoben werden. Eine andere Möglichkeit besteht darin, von den Content-Anbietern für die Aufnahme in die Suchmaschine eine *Anmeldegebühr* zu verlangen, eine weitere Variante ist die *Platzierungsgebühr*. Damit ist gemeint, dass die Content-Anbieter dafür bezahlen, wenn ihre Angebote bei bestimmten Suchbegriffen an vorderer Stelle platziert werden. Gleichgültig für welche dieser Varianten sich ein Suchmaschinenbetreiber entscheidet, in jedem Fall ist es für ihn – zumindest mittelfristig – von zentraler Bedeutung, dass möglichst viele User möglichst häufig und lange die Suchmaschine nutzen (vgl. Gartz 2000: 43). Um dieses Ziel zu erreichen, müssen die Anbieter sicherstellen, dass die User mit der Qualität der Leistung zufrieden sind, denn nur dann werden sie die Maschine wieder benutzen.

Aus der Sicht der *User* muss eine Suchmaschine vor allem eins tun: Sie muss als Antwort auf eine Suchanfrage (mög-

lichst viele) Web-Seiten auflisten, auf denen die *gesuchten* Informationen zu finden sind. Darüber hinaus sollte die Anfrage möglichst einfach durchzuführen sein, gleichzeitig sollte es aber die Option geben, komplexe Suchbefehle zu formulieren. Schließlich sollten die Ergebnisse möglichst schnell und übersichtlich präsentiert werden (vgl. Potempa et al. 1998: 130 ff.; Hartmann, Näf & Schäuble 2000: 134 ff.). Das wichtigste Qualitätskriterium aus der Sicht des Users ist aber sicherlich, dass relevante und möglichst zentrale Informationen zu den entsprechenden Anfragen präsentiert werden. Die Relevanz der Ergebnisse ist dementsprechend auch bei den Verbrauchertests von Suchmaschinen immer das wichtigste Kriterium der Qualitätsbeurteilung (vgl. Stiftung Warentest 2001; Becker 2002).

Die drei beteiligten Akteure haben demnach unterschiedliche Ziele. Für die Suchmaschinenbetreiber ist es allerdings notwendig, dass sie sich an den Interessen der User orientieren, wenn sie ihre eigenen ökonomischen Ziele dauerhaft erreichen wollen. Das heißt, sie müssen dafür sorgen, dass die User zufrieden sind. Dies geht jedoch nur, indem sie ihnen helfen, geeignete Informationen zu den gesuchten Themen zu finden. Wenn die User mit den Ergebnissen ihrer Suchanfragen unzufrieden sind, werden sie zu anderen Suchmaschinen wechseln, was für die Existenz des Angebots gefährlich wäre. Gleichgültig welche Finanzierungsform die Betreiber einer Suchmaschine gewählt haben, wenn die Maschine nicht im ausreichenden Maß genutzt wird, kann das Angebot wirtschaftlich nicht bestehen. Die Interessen der Suchmaschinenbetreiber sind auf diese Weise mit den Interessen der User verknüpft. Die Interessen der Content-Anbieter sind hingegen anders: Ihnen geht es *nicht* darum, dass die User die *besten* Informationen finden, sondern sie wollen, dass die User *ihre* Informationen finden, gleichgültig ob diese für den User auch die «besten» sind oder nicht.

#### **Die Suchdienstvarianten:**

##### **Robots, Kataloge, Metasuchmaschinen**

Für die nachfolgenden Ausführungen ist es notwendig zu definieren, welche Dienste im WWW gemeint sind, wenn von «Suchmaschinen» die Rede ist.<sup>3</sup> Der Begriff «Suchmaschine» wird als Sammelbegriff für verschiedene Angebotsformen verwendet, die eine gezielte Suche nach Informationen ermöglichen, und zwar nach solchen Informationen, von denen die Nutzer nicht genau wissen, wo sie im Internet zu finden sind. Der Begriff erweckt den Eindruck, dass es sich dabei um Maschinen handelt, die tatsächlich eigenständig nach Informationen suchen. Das trifft aber nicht auf alle Angebote zu. Man kann drei unterschiedliche Idealtypen von Suchmaschinen unterscheiden: *Robots* (auch Crawler, Spider oder Worms genannt), *Kataloge* (auch Webverzeichnisse genannt) und *Metasuchmaschinen*. In der Praxis finden sich aber auch Mischtypen, sogenannte hybride Formen (vgl. Gartz 2000: 47), und die Übergänge zwischen den verschiedenen Formen sind fließend.<sup>4</sup>

*Kataloge* entsprechen am wenigsten dem, was man mit dem Namen Suchmaschine assoziiert. Ein Katalog ist ein – gewöhnlich hierarchisch geordnetes – Verzeichnis von Web-Seiten (vgl. Potempa et al. 1998: 83 ff.; Kuhlen 1999: 237 ff.). Die Aufnahme der Web-Seiten in den Katalog erfolgt meistens dadurch, dass die Content-Anbieter ihre Seiten beim jeweiligen Katalog anmelden. Eine aktive Suche nach Webangeboten von Seiten der Katalogbetreiber erfolgt in der Regel nicht. Die bei dem Verzeichnis zur Anmeldung vorgelegten Seiten werden von Redakteuren des Katalogs begutachtet, und je nachdem, ob sie die vorgeschlagenen Web-Angebote für geeignet (qualitativ hochwertig) erachten, werden die Seiten in den Katalog aufgenommen oder abgelehnt. Angebote, die nicht aufgenommen werden, können später von den Usern natürlich nicht im Katalog gefunden werden. Hier ist es offensichtlich, dass die Redakteure der Kataloge in geradezu klassischer Weise die Funktion von Gatekeepern erfüllen. Sie bestimmen darüber, welche Informationen überhaupt für die Nutzer zur Verfügung gestellt werden und an welcher Stelle des Verzeichnisses sie platziert werden. Teilweise werden die Angebote auch von den Redakteuren kommentiert, ebenfalls eine typisch journalistische Tätigkeit. Nach welchen Kriterien die Redakteure dabei entscheiden, über welche fachlichen oder journalistischen Kompetenzen sie verfügen und unter welchen Bedingungen sie arbeiten, ist aus wissenschaftlicher Perspektive weitgehend unerforscht. Qualifikation und Arbeitsbedingungen haben aber zweifellos einen großen Einfluss auf die Qualität des Angebots (vgl. Philippus 1997: 69). Kataloge sind durchaus kein randständiges Mittel der Informationssuche. Bei einigen der bekanntesten Suchdienste (Yahoo, Web.de, MSN) handelt es sich um Kataloge (vgl. ITXPLAIN). Die Tatsache, dass es sich bei den Retrievals solcher Suchmaschinen um das Ergebnis redaktioneller Entscheidungen handelt, dürfte vielen Nutzern nicht bewusst sein.<sup>5</sup>

Die *Robots* unterscheiden sich von den Katalogen in mehrfacher Hinsicht. Der wichtigste Unterschied besteht darin, dass Menschen bei diesen Suchmaschinen nicht unmittelbar in die Auswahl und Klassifizierung der Web-Seiten involviert sind. Diese Funktion übernehmen Programm-routinen. Ein weiterer Unterschied besteht in der aktiven Suche: Im Gegensatz zu Katalogen reagieren *Robots* nicht nur auf Seiten, die zur Anmeldung vorgelegt werden, sondern suchen auch aktiv nach Web-Seiten und nehmen diese in den Datenbestand auf.

Aufgrund der Größe des Internets kann kein Robot die weltweit vorhandenen Web-Seiten vollständig erfassen. Das bedeutet, dass jeder Nutzer, der nur *einen* Suchdienst verwendet, auf seine Suchanfrage nicht eine Auflistung *aller* relevanter Seiten im Netz erhält, sondern nur die eines relativ kleinen *Ausschnitts* (vgl. Loosen 1999: 43, Gartz 2000: 44 f.). Dieses Problem können *Metasuchmaschinen* bis

zu einem gewissen Grad beheben. Metasuchmaschinen suchen nicht selbst nach Informationen im Netz und sie legen auch keine Kataloge an, sondern sie leiten eine an sie gerichtete Suchanfrage gleichzeitig an verschiedene *Robots* und Kataloge weiter. Die Ergebnisse dieser verschiedenen Suchanfragen werden dann – je nach Metasuchmaschine im unterschiedlichen Maße – verarbeitet und dem User wird ein fusioniertes Ergebnis präsentiert (vgl. @-web Webmagazin).<sup>6</sup> Metasuchmaschinen können Probleme, die durch die begrenzte Kapazität der einzelnen Maschinen entstehen, teilweise lösen, allerdings wird das Problem der Relevanzzuweisung an einzelne Dokumente bei den Metasuchmaschinen noch komplexer und problematischer (vgl. Hartmann, Näf & Schäuble 2000: 74 f.).

#### **Die Arbeitsweise der Robots:**

##### **Suchen, auswählen, indizieren, ranken**

Im Folgenden soll vor allem die Arbeitsweise der *Robots* näher erläutert werden, denn sie sind die am häufigsten genutzten Suchmaschinen. Auf die *Robots* bezieht sich auch das eingangs wiedergegebene Zitat zum «Wettrüsten». Vor allem am Beispiel der *Robots* muss geprüft werden, ob sich die These von den Suchmaschinen als «Gatekeeper des Internets» aufrecht erhalten lässt.

Was machen *Robots*? Bei den *Robots* selbst handelt es sich um eine *Suchsoftware*. *Robots* suchen im Internet nach Web-Seiten, fragen deren Angebote ab und verfolgen dann die Links, die auf den Web-Seiten enthalten sind.<sup>7</sup> Die Suchsoftware leitet die auf den Web-Seiten gefundenen Informationen anschließend an eine Indizierungssoftware weiter. Diese wandelt die Informationen der Web-Seiten in einen Index um, der schließlich in einer Datenbank gespeichert wird. Wenn ein User ein Suchanfrage stellt, werden die Angaben dieser Suchanfrage mit Hilfe einer Retrievalsoftware mit den indizierten Daten der Datenbank verglichen, passende Einträge werden ermittelt, sortiert und in einer Trefferliste ausgegeben (vgl. Potempa et al. 1998: 95 ff.; @-web Webmagazin).

Suchmaschinen suchen also nicht nur nach Informationen, sondern sie wandeln diese Informationen anschließend in Indizes um. Dabei «entscheiden» sie, welche Informationen eine Seite enthält und welche Relevanz diese Informationen haben.<sup>8</sup> Bei einer Anfrage durch den User bringen sie schließlich die ermittelten und indizierten Web-Seiten in eine Rangreihe. Bei dieser Reihung handelt es sich zwar nicht um einen echten Selektionsprozess, aber der Effekt einer solchen sortierten Reihe ist mit einer Selektionsentscheidung identisch. Web-Seiten, die nicht auf den vorderen Plätzen einer Retrievalliste aufgeführt werden, stehen den Usern zwar theoretisch zur Verfügung, faktisch werden jedoch fast immer nur die ersten zehn bis zwanzig Einträge von den Usern verwendet. Neben der Suchphase muss also vor allem der Indizierungs- und Rankingpro-

zess daraufhin untersucht werden, ob dort Selektions- bzw. Gatekeepingprozesse stattfinden und wie diese organisiert sind. Aber schon an dieser Stelle kann zweifelsfrei festgehalten werden: Das Ergebnis dieses Prozesses ist bei fast allen Suchanfragen eine ungeheure Selektion, denn in den meisten Fällen sind die gesuchten Begriffe in Tausenden von Web-Dokumenten zu finden. Davon können dann jedoch nur einige wenige ausgewählt werden, indem sie auf den vorderen Plätze der Retrievallistings positioniert werden.

In der *Suchphase* entscheidet sich, welche Web-Seiten überhaupt von den Robots gefunden werden. Keine Suchmaschine der Welt ist in der Lage, das gesamte Web zu durchsuchen. Es stellt sich also die Frage, wie eine Suchmaschine entscheidet, *wo* sie im Netz sucht. Zwei Wege sind dabei vor allem von Bedeutung: Zum einen werden Web-Seiten von Content-Anbietern bei den Suchmaschinen angemeldet.<sup>9</sup> Diese angemeldeten Web-Seiten werden dann in einem überschaubaren Zeitraum von den Suchmaschinen aufgesucht und indiziert. Zum anderen verfolgen die Robots die Links auf den angemeldeten und bereits indizierten Seiten und gelangen so auf weitere Angebote und indizieren diese ebenfalls (vgl. @-web Webmagazin). Ob eine Web-Seite von einem Robot gefunden wird, ist also im hohen Maß davon abhängig, ob sie angemeldet wurde. Darüber hinaus spielt auch die Frage, ob sie mit anderen Web-Seiten verlinkt ist, eine Rolle. Isolierte Seiten, die weder verlinkt noch angemeldet wurden, können durch Robots nicht gefunden werden (Hartmann, Näf & Schäuble 2000: 64). Soweit bekannt, werden bislang fast alle von den Robots gefundenen Web-Seiten auch indiziert – zumindest von den universellen, nicht auf einzelne Themen oder Länder spezialisierten Maschinen.<sup>10</sup> Allerdings gibt es bereits Ansätze, bestimmte Angebote schon auf dieser Ebene – also noch vor der Indizierung – auszusondern. Dazu gehören beispielsweise pornografische, gewaltverherrlichende oder rechtsradikale Web-Angebote (vgl. Hartmann, Näf & Schäuble 2000: 73), aber mittlerweile auch ganze Domains. Vor allem Anbieter kostenloser Homepages haben offenbar das Problem, dass die Seiten, die auf diesen Domains positioniert werden, nicht von den Suchmaschinen indiziert werden (vgl. V-business: Zukunft der Suchmaschinen). Der Einsatz solcher Selektionsroutinen könnte zum systematischen Ausschluss ganzer Angebotsformen und Inhalte führen. Eine solche Vorgehensweise ist aus *gesellschaftlicher* Perspektive nicht unproblematisch, da dadurch die Gefahr besteht, dass das egalitäre und demokratische Potential des Internets erheblich beeinträchtigt wird.<sup>11</sup> Bedenklich ist vor allem, wenn die User nicht darüber informiert werden, nach welchen Kriterien einzelne Inhalte oder Domains ausgegrenzt werden. Trotz dieser Tendenzen haben die Robots in der Suchphase – zumindest bislang – nur sehr begrenzt die Funktion von Gatekeepern. Hier sind es primär die Content-Anbieter selbst, die durch ihre Initia-

tive (die Anmeldung bei der Suchmaschine) dafür sorgen müssen, dass ihre Angebote gefunden werden können. Die prinzipielle Auffindbarkeit von Informationen hängt also vor allem von der «Öffentlichkeitsarbeit» der Anbieter ab.

Anders sieht es in der *Phase der Indexbildung* aus: Nachdem eine Web-Seite von einer Suchmaschine gefunden worden ist, wird ihr Inhalt an die Indizierungssoftware weitergeleitet. Die Indizierungssoftware ermittelt, welche Inhalte die Web-Seite enthält und «entscheidet», wie zentral und relevant die einzelnen Begriffe für den Inhalt dieser Seite sind bzw. wie gut der Inhalt einer Web-Seite durch bestimmte Begriffe repräsentiert wird.<sup>12</sup> Indirekt wird damit schon zu diesem Zeitpunkt festgelegt, welche Themenrelevanz eine Web-Seite im Vergleich zu anderen Seiten hat. Um diese «Entscheidungen» treffen zu können, kann die Software auf verschiedene Informationen zurückgreifen, die auf den Web-Seiten enthalten sind. Dabei unterscheiden sich die verschiedenen Suchmaschinen darin, welche dieser Informationen sie für die Gewichtung heranziehen und welche Bedeutung die jeweiligen Informationen für die Indizierung und das Ranking haben. Die Zuschreibung von Relevanz erfolgt also bei der Indizierung. Die in dieser Phase zugeschriebene Wichtigkeit wirkt sich später auf die Position der indizierten Web-Seite in der Retrievalliste aus, wenn eine entsprechende Suchanfrage durch einen User erfolgt.<sup>13</sup>

#### **Die Kriterien der Relevanzzuschreibung:**

##### **Primäre und sekundäre Kriterien**

Folgende Kriterien können für die Relevanzzuschreibung bei der Indizierung und damit für das Retrievalranking von Bedeutung sein: Wichtig ist zunächst einmal die Position, an der sich das gesuchte Wort auf der Web-Seite befindet. Wenn sich ein Suchwort am Anfang einer Web-Seite befindet, dann wird es als bedeutsamer für die Beschreibung des Inhalts angesehen als wenn es weiter unten auf der Seite zu finden ist. Wenn der gesuchte Begriff in der Überschrift auftritt oder wenn er in irgendeiner Form gestalterisch hervorgehoben wurde (fett, größere Schrift etc.), kann dies ebenfalls vom Indizierungsprogramm als Indikator für erhöhte Relevanz interpretiert werden. Einige Suchmaschinen analysieren auch die URL, d.h. den Pfad und den Dateinamen. Wenn der Suchbegriff im Namen oder im Pfad einer Web-Seite vorkommt, dann wird dies als Relevanzkriterium bei der Indizierung berücksichtigt (vgl. Gartz 2000: 150 ff.). Bei Hybrid-Suchmaschinen kann auch berücksichtigt werden, ob eine Seite bereits katalogisiert wurde. Solche inhaltlich geprüften Seiten können dann beim Retrieval bevorzugt werden (vgl. Potempa et al. 1998: 125).

Neben der Position und den gestalterischen Merkmalen wird vor allem die *Häufigkeit* der Begriffs-nennung als Indikator herangezogen.<sup>14</sup> Je häufiger der gesuchte Begriff

in einem Dokument vorkommt, desto höher ist die Wahrscheinlichkeit – so die Logik einiger Suchmaschinen – dass es sich bei der Web-Seite um relevanten Content für den User, der diesen Suchbegriff eingegeben hat, handelt. Problematisch an einem solchen Algorithmus ist allerdings, dass Anbieter, die sich um sprachliche Abwechslung und Vielfalt bemühen, indem sie Synonyme oder Umschreibungen für die zentralen Begriffe verwenden, dadurch beim Ranking benachteiligt werden können.<sup>15</sup>

Hinsichtlich der Begriffsfrequenz spielt nicht nur die absolute Häufigkeit, sondern auch die *relative Häufigkeit* eine Rolle. Das Relevanzkriterium «relative Häufigkeit» besagt: Wenn ein Suchbegriff in einem kurzen Dokument häufig auftritt, dann hat dieses Dokument eine höhere Themenrelevanz als wenn der Begriff in gleicher Häufigkeit in einem langen Dokument vorkommt. Das hat zur Folge, dass kurze Seiten tendenziell prominenter gelistet werden als umfangreiche Dokumente (vgl. drweb.de 2002). Dadurch besteht allerdings die Gefahr, dass ausführliche Informationen und Hintergrundberichte nicht auf den vorderen Plätzen einer Retrievalliste erscheinen.

Darüber hinaus spielen auch die sogenannten *Meta-Tags* eine gewisse Rolle. Meta-Tags sind unsichtbare Bestandteile der HTML-Datei. Sie können aber von den Robots gelesen und interpretiert werden. Einige dieser Meta-Tags sind explizit dafür vorgesehen, den Anbietern der Web-Seiten die Möglichkeit zu geben, die Inhalte ihrer Seiten zu beschreiben und dadurch die Suchmaschinen bei der inhaltlichen Klassifizierung der Seiten zu «unterstützen». Von Bedeutung für die Indizierung sind vor allem die Meta-Tags «Keywords», «Description» und «Abstract». Wenn ein Suchbegriff in den Meta-Tags vorkommt, dann bringt der Content-Anbieter damit zum Ausdruck, dass auf dieser Web-Seite relevante Inhalte zu diesem Begriff zu finden sind (vgl. Gartz 2000: 151 f.).

Neben diesen *primären Kriterien*, die sich auf den Inhalt der Web-Seite beziehen, gibt es aber auch *sekundäre Relevanzattribute*, die aus der Struktur des Angebots abgeleitet werden. Sie sind deswegen als sekundäre Attribute zu bezeichnen, weil sie nur dann wirksam werden, wenn zuvor durch die primären Attribute eine inhaltliche Relevanz festgestellt worden ist. Zu den sekundären Attributen gehören die Link-Relevanz und die Link-Popularität. Eine Web-Seite hat dann eine hohe Link-Relevanz, wenn von dieser Seite viele Links auf andere URLs verweisen, in denen ebenfalls der gesuchte Begriff vorkommt. Mit Link-Popularität wird die entgegengesetzte Verweisstruktur bezeichnet: Die Link-Popularität einer Web-Seite ist hoch, wenn von vielen thematisch einschlägigen Seiten auf sie verwiesen wird. Seiten mit hoher Link-Relevanz und hoher Link-Popularität werden von manchen Suchmaschinen als besonders relevant eingestuft (vgl. Hartmann, Näf & Schäuble 2000: 45 f. & 55).

Weitere *strukturelle Relevanzkriterien* sind die Größe (Seitenzahl) einer Domain und die Directory-Tiefe. Web-Seiten, die zu einer großen Domain gehören und/oder Teil eines tiefgestaffelten Verzeichnissystems sind, wird eine größere Relevanz zugeschrieben als Web-Seiten aus kleinen einfachstrukturierten Angeboten. Thematisch homogene Domains werden ebenfalls von einigen Indizierungs- bzw. Retrievalalgorithmen bevorzugt (vgl. Explido).

Eine weitere Dimension sekundärer Relevanzzuschreibung ist die *Zeitkomponente*: Zum einen wird Web-Seiten, die bereits etabliert sind, d.h. deren URL sich bereits lange im Datenbestand der Suchmaschine befindet, eine höhere Bedeutsamkeit zugeschrieben als neuen Angeboten; andererseits werden Seiten, die erst vor kurzer Zeit aktualisiert wurden, bei der Indizierung als relevanter eingestuft als solche, die schon lange nicht mehr verändert worden sind.

Schließlich wird auch die *Nutzungsfrequenz* der ermittelten Web-Seiten von einigen Suchmaschinen als Relevanzattributor für die nachfolgenden Suchanfragen verwendet. Web-Seiten, die nach einer bestimmten Suchanfrage von den Usern besonders häufig aufgerufen werden, wird eine höhere Ergebnisrelevanz zugeschrieben und sie werden bei späteren Suchanfragen mit dem gleichen Suchbegriff (noch) weiter vorne auf der Retrievalliste platziert.<sup>16</sup> Durch diesen Algorithmus wird die Selektivität permanent durch Rückkopplung verstärkt (vgl. Loosen 1999: 45).

#### **Die «geheimen»Kriterien:**

##### **Das Wettrüsten und die Folgen**

Die Kriterien, nach denen die Seiten indiziert und gerankt werden, variieren zwischen den verschiedenen Maschinen. Je nachdem, welche Relevanzindikatoren von der jeweiligen Suchmaschine herangezogen werden, unterscheiden sich die Retrievals mehr oder weniger deutlich. Die Indizierungsroutinen und Rankingalgorithmen haben Einfluss darauf, welche Seiten an welcher Position platziert werden. Von daher haben sie die Funktion von Gatekeepern: Sie beeinflussen, was die User präsentiert bekommen und was nicht.<sup>17</sup> Die genauen Regeln, nach denen eine Suchmaschine die Web-Seiten indiziert und rankt, werden von den Suchmaschinenbetreibern nicht bekannt gegeben. Hierfür gibt es zwei Gründe: Zum einen sind die Rankingkriterien das «Erfolgsgeheimnis» einer Suchmaschine, die natürlich gegenüber der Konkurrenz nicht offengelegt werden sollen (vgl. Hartmann, Näf & Schäuble 2000: 44). Wichtiger noch als die Geheimhaltung gegenüber der Konkurrenz ist jedoch die Geheimhaltung der Kriterien gegenüber den Content-Anbietern, denn genau um diese Kriterien geht es beim «Wettrüsten» zwischen den Suchmaschinenbetreibern und den Content-Anbietern. Die Content-Anbieter (bzw. die darauf spezialisierten Dienstleister) versuchen die Indizierungskriterien der Suchmaschinen zu ermitteln und die Rankingalgorithmen zu entschlüsseln, um

daraufhin ihre Content-Angebote dahingehend zu optimieren, dass sie – bei entsprechenden Anfragen – von den Suchmaschinen auf den vorderen Rankingplätzen gelistet werden.<sup>18</sup> Durch eine solche Anpassung werden die Kriterien aber mit der Zeit in ihrer Wirksamkeit eingeschränkt: Die Suchmaschine schreibt dann nicht mehr den wirklich relevanten Seiten eine hohe Relevanz zu, sondern den Web-Seiten, deren Betreiber am besten in der Lage waren, ihr Angebot an die Indizierungs- und Rankingkriterien anzupassen und dahingehend zu optimieren. Eine solche Strategie erinnert an die Vorgehensweise von Öffentlichkeitsarbeitern, die die Selektionsroutinen der Journalisten (z.B. in Form von Nachrichtenfaktoren) benutzen, um nachrichtenwerte Berichte zu produzieren, welche die Journalisten dann – weil sie genau ihren Auswahlkriterien entsprechen – direkt übernehmen können.

Diese Anpassung der Web-Seiten an die Indizierungskriterien der Software hat zur Folge, dass die Suchmaschinenbetreiber ihre Kriterien immer wieder umstellen bzw. neue Kriterien hinzufügen müssen, um sich gegenüber den «Manipulationsversuchen» der Anbieter zur Wehr zu setzen. So wird beispielsweise die absolute oder relative Häufigkeit der Verwendung eines bestimmten Suchworts von einigen Suchmaschinen nur noch eingeschränkt (oder gar nicht) als Relevanzindikator angesehen. Wenn ein Begriff auf einer Web-Seite zu häufig auftritt, dann wird dies vom Indizierungsalgorithmus als «Manipulationsversuch» (Spamming) interpretiert und die Seite wird beim Listing nicht berücksichtigt. Ein anderes Beispiel aus diesem Problembereich betrifft den Einsatz von Meta-Tags: Wenn zwischen den Inhalten, die durch die Meta-Tags beschrieben werden und den Inhalten der sichtbaren Web-Seite deutliche Diskrepanzen bestehen, wird dies von einigen Maschinen ebenfalls als Spamming gewertet und die betreffende Seite wird ignoriert und nicht indiziert (vgl. Hartmann, Näf & Schäuble 2000: 47).

Dieser Konflikt zwischen Suchmaschinenbetreibern und Content-Anbietern ist für die Kommunikation im Internet nicht folgenlos: Vor allem, wenn man die oben aufgeführten *sekundären* Relevanzkriterien betrachtet, kommt man zu der Einschätzung, dass deren Anwendung durchaus problematische Folgen haben kann, denn die strukturellen Kriterien bedrohen tendenziell das offene, hierarchiefreie Nebeneinander unterschiedlicher Netzangebote. Wenn nicht mehr allein der *Inhalt* darüber entscheidet, ob dem User ein bestimmtes Web-Angebot auf vorderen Rankingplätzen präsentiert wird, sondern die Informationen darüber an Bedeutung gewinnen, *wo* diese Web-Seite zu finden ist und wie sie in ein *Zitiergefüge* (durch Links) eingebunden ist, dann stellt sich die Frage, welche Folgen solche Indizierungs- und Rankingmechanismen haben. Aus *gesellschaftlicher* Perspektive ist eine solche Auswahlstrategie vor allem deswegen problematisch, weil durch diese Logik

große, etablierte Anbieter systematisch bevorzugt werden. Wenn darüber hinaus auch noch die Nutzungszahlen als zusätzlicher Relevanzverstärker verwendet werden, dann haben kleinere Anbieter mittelfristig keine Chancen, überhaupt noch von den Usern gefunden und wahrgenommen zu werden.<sup>19</sup>

Nicht nur durch diese beschriebenen Folgen des «Wettrüstens», auch von anderer Seite wird die Unabhängigkeit der von den Suchmaschinen erbrachten Kommunikationsleistungen bedroht: Jeder Anbieter einer Suchmaschine steht vor der Frage, wie er sein Dienstleistungsangebot finanzieren soll. Die verschiedenen Finanzierungsvarianten wurden bereits oben kurz genannt. Eine kostendeckende Finanzierung der Suchmaschinen allein durch Banner-Werbung wird zunehmend schwieriger. Die Alternativen dazu sind jedoch aus *gesellschaftlicher* Sicht wenig überzeugend. Wenn die Tendenz dahin gehen sollte, dass die Aufnahme von Web-Seiten in den Index der Suchmaschinen nur noch gegen Gebühr erfolgt, dann werden die Suchmaschinen mittelfristig zu Branchenbüchern, denn nur die kommerziellen, finanzkräftigen Anbieter werden bereit und in der Lage sein, für die Aufnahme in die Suchmaschinen zu bezahlen. Wenn hingegen vordere Rankingplätze direkt an zahlende Anbieter verkauft werden, in Auktionen versteigert werden oder das Ranking durch Sponsorinteressen beeinflusst wird (vgl. Loosen 1999: 46; Potempa et al. 1998: 125), dann bedeutet dies, dass Informationsanfragen der User mit Werbung beantwortet werden. Eine klare Trennung von redaktionellem Angebot und Werbung – wie sie von den traditionellen Medien selbstverständlich verlangt wird – ist dann im Bereich der Informationssuche mit Hilfe von Suchmaschinen vollkommen hinfällig. Dies ist zweifellos ebenfalls keine wünschenswerte Perspektive. Aber auch die letzte der oben genannten Varianten ist nicht unproblematisch: Wenn die Nutzung der Suchmaschinen für die User kostenpflichtig würde, dann könnte dies zur Folge haben, dass dadurch eine neue Barriere errichtet wird, die einer chancengleichen Nutzung des Internets durch die verschiedenen Bevölkerungssegmente entgegensteht.

### Resümee

Die Relevanz von Suchmaschinen für die Kommunikationsprozesse im Internet ist kaum zu überschätzen. Die strategische Bedeutung von Suchmaschinen hat dazu geführt, dass zwischen Content-Anbietern und Suchmaschinenbetreibern ein Kampf ausgetragen wird, bei dem es letztendlich um die Verfügungshoheit über die Auswahl und Präsentation von Informationen geht. In diese Auseinandersetzung sind zur Zeit vermutlich vor allem die kommerziellen Anbieter verwickelt. Aber sie betrifft letztendlich alle Beteiligten, denn die oben dargestellten Konsequenzen dieses Konflikts haben Auswirkungen in allen Themenfeldern auf die Auswahl von Web-Seiten und deren Positionierung. Und mittelfristig können die Conse-

quenzen noch deutlich darüber hinaus gehen: Wenn die Zahl der verfügbaren Informationen im Netz wächst, wird die Bedeutung der Suchmaschinen in ihren unterschiedlichen Varianten weiter zunehmen, auch für Informationsbereiche, die heute noch nicht schwerpunktmäßig im Netz angesiedelt sind. Menschliche Redakteure werden dann – nicht zuletzt bedingt durch den Kostenfaktor – auch in klassischen Informationsbereichen zunehmend durch «intelligente» Softwarelösungen ersetzt werden. Dies ist sicherlich eine, sowohl in Bezug auf den Arbeitsmarkt als auch in gesellschaftlicher Hinsicht, brisante Perspektive (vgl. Hack 1998: 200 f.).

Einige der möglichen problematischen Folgen wurden oben dargestellt. In wie weit bereits heute bestimmte Angebote des Internets durch die Logik der Suchmaschinen bei der Indizierung systematisch vernachlässigt werden – weil ihnen auf Grund struktureller Kriterien eine «zu» geringe Relevanz zugebilligt wird – kann hier nicht beantwortet werden. Um diese Frage zu beantworten, bedarf es empirischer Forschung. Vor allem experimentelle Untersuchungen müssten dazu durchgeführt werden, bei denen beispielsweise einzelne der oben diskutierten Relevanzattributoren gezielt variiert werden, um zu überprüfen, welche Auswirkungen diese Faktoren auf die Retrievallisten haben. Solche Untersuchungen sind mit einem erheblichen Aufwand verbunden und müssen zudem kontinuierlich wiederholt werden, da auch die Auswahl- und Positionierungskriterien permanent verändert werden. Im Bereich der kommerziellen Forschung werden Untersuchungen dieser Art bereits seit langem durchgeführt, da dort die Relevanz der Suchmaschinen für den Geschäftserfolg schon frühzeitig erkannt wurde. Auch die oben erwähnten Verbrauchertests sind diesem Untersuchungsfeld zuzuordnen. Allerdings ist das Erkenntnisinteresse sowohl der marketingorientierten als auch der verbraucherorientierten Untersuchungen anders ausgerichtet. Mit den oben dargestellten Überlegungen sollte gezeigt werden, dass es neben diesen beiden Perspektiven auch eine publizistische Betrachtungsweise gibt, die die Kommunikationsleistungen der Suchmaschinen unter *gesellschaftlichen* Gesichtspunkten betrachtet und diskutiert. Es wird Zeit, die Gatekeeper im Internet auch aus dieser Perspektive genauer zu analysieren und das Bewusstsein dafür zu schärfen, dass die Kommunikationsleistungen der Suchmaschinen durchaus nicht neutral sind.

Die technische Entwicklung wird vermutlich dahin gehen, dass Softwareagenten – sogenannte «intelligente Suchmaschinen» (vgl. Potempa et al. 1998: 214 ff.) – auf individuelle Bedürfnisse zugeschnittene Informationsleistungen erbringen, indem sie zunächst – wie klassische Suchmaschinen – Informationen beschaffen, dann aber im Push-Modus (vgl. Kuhlen 1999: 258 ff.; Hartmann, Näf & Schäuble 2000: 125 ff.) diese individualisiert den Usern

zusenden. Spätestens dann begibt sich der Einzelne vollständig in die Hand der automatisierten Kommunikationsdienstleister. Nicht zuletzt angesichts dieser Perspektive ist die Frage zu stellen, ob und in wie weit Vertrauen in die durch Suchmaschinen erbrachten Kommunikationsleistungen begründet ist, oder ob nicht ein Konzept des «konstruktiven Misstrauens» angemessener wäre (vgl. Kuhlen 1999: 124 ff.). Eine gesellschaftliche und wissenschaftliche Auseinandersetzung mit diesem Thema, bei der auch die Gefahr bewusster Manipulation und Zensur (vgl. Hack 1998: 190 ff.) nicht ausklammert werden sollte, erscheint dringend notwendig.

**Dr. Jens Wolling**, Studium der Kommunikationswissenschaft an der FU Berlin, ist seit 1999 wissenschaftlicher Assistent im Fachgebiet Politikwissenschaft/Medien des Instituts für Medien- und Kommunikationswissenschaft der TU Ilmenau.

*jens.wolling@tu-illmenau.de*

#### Literatur

- @-web Webmagazin (ohne Jahr): Suchmaschinen.  
URL: <http://www.at-web.de> [Download 13.02.02].
- ARD-/ZDF-Online-Studie 2001 (unveröffentlichter Tabellenband).
- Baerns, Barbara (1987): Macht der Öffentlichkeitsarbeit und Macht der Medien. In: Sarcinelli, Ulrich (Hrsg.): Politikvermittlung. Bonn, S. 147–160.
- Becker, Alexander (2002): 10 Suchmaschinen im Test. In: Tomorrow 3/2002, S. 26–29.
- Bekavac, Bernard (2001): Tutorial zur Suche im WWW/Internet, Methoden und Verfahren von Suchdiensten im WWW/Internet.  
URL: <http://www.inf-wiss.uni-konstanz.de/suche/> [Download 13.02.02].
- Die Suchfibel (ohne Jahr): URL: <http://www.suchfibel.de>
- Donsbach, Wolfgang (1997): Einleitung. Legitimität und Effizienz von PR. (Hrsg.): In: Derselb.: Public Relations in Theorie und Praxis. München, Reinhard Fischer.
- drweb.de online magazin (2001): Suchmaschinen Ranking: Worauf kommt es eigentlich an?  
URL: <http://www.ideenreich.com/suchmaschinen/ranking.shtml> [Download 13.02.02].
- drweb.de online magazin (2002): Suchmaschinen - Die Themen.  
URL: <http://www.ideenreich.com/suchmaschinen/index.shtml> [Download 13.02.02].
- Eimeren, Birgit van, Heinz Gerhard & Beate Frees (2001): ARD/ZDF-Online-Studie 2001: Internetnutzung stark zweckgebunden. In: Media Perspektiven (8), S. 382–397.
- Explido (ohne Jahr): [http://www.promotionwelt.de/suchmaschinen\\_optimierung/suchmaschinen\\_rankingverfahren.html](http://www.promotionwelt.de/suchmaschinen_optimierung/suchmaschinen_rankingverfahren.html) [Download 13.02.02].
- GAB + GESSINGER (2001): Web-Ranking. Stärkung des Online-Auftrittes durch bessere Positionierung in Suchmaschinen und Katalogen.  
URL: <http://www.gab-gessinger.de/de/web-ranking-2001-08.pdf> [Download 13.02.02].
- Gartz, Joachim (2000): Professionelle Internetrecherche für Wissenschaftler und Online-Profis. Kilchberg, Smartbooks Publishing.
- Hack, Günter (1998): Secret Services. Künstliche Intelligenzen auf Datenjagd im Netzwerk. In: Glotz, Peter (Hrsg.): Die Benach-

richtung der Deutschen: aktuelle Fernsehberichterstattung zwischen Quoten- und Zeitzwang. Frankfurt am Main, Institut für Medienentwicklung und Kommunikation (IMK) in der Verlagsgruppe FAZ, S. 176–202.

- Hartmann, Werner, Michael Näf & Peter Schäuble (2000): Informationsbeschaffung im Internet: Grundlegende Konzepte verstehen und umsetzen. Zürich, Orell Füssli.
- ITXPLAIN Internet-Systemhaus und Beratungsgesellschaft mbH (ohne Jahr): Suchmaschinen verstehen. URL: <http://www.suchmaschinen-verstehen.de/> [Download 13.02.02].
- Kubicek, Herbert & Stefan Welling (2000): Vor einer digitalen Spaltung in Deutschland? Annäherung an ein verdecktes Problem von wirtschafts- und gesellschaftspolitischer Brisanz. In: *Medien & Kommunikationswissenschaft* 48 (4), S. 497–517.
- Kuhlen, Rainer (1999): Die Konsequenzen von Informationsassistenten: Was bedeutet informationelle Autonomie oder wie kann Vertrauen in elektronische Dienste in offenen Informationsmärkten gesichert werden? Frankfurt am Main, Suhrkamp.
- Loosen, Wiebke (1999): Suchmaschinen. «Informations- und Wissensverwalter» im World Wide Web. In: *Medien Journal* (3), S. 42–48.
- Philippus, Thomas (1997): Informationssuche im Internet: Tips für Profis. Berlin Offenbach, VDE.
- Potempa, Thomas, Peter Franke, Wilfried Osowski & Maria-Elisabeth Schmidt (1998): *Informationen finden im Internet*. München Wien, Carl Hanser.
- Searchenginewatch (ohne Jahr) URL: <http://www.searchenginewatch.com>
- Shoemaker, Pamela J. (1991): *Gatekeeping*. Newbury Park California.
- Stiftung Warentest (2001): Internet-Suchmaschinen. Google trifft am besten. URL: [http://www.warentest.de/pls/sw/SW.Main?p\\_KNr=5001211190948920011023145541&p\\_E1=1&p\\_E3=50&p\\_E4=30&p\\_id=21394](http://www.warentest.de/pls/sw/SW.Main?p_KNr=5001211190948920011023145541&p_E1=1&p_E3=50&p_E4=30&p_id=21394) [Download 13.02.02].
- V-business (ohne Jahr): Nutzung der Suchmaschinen. URL: <http://www.suchmaschinen-marketing.de/nutzungsuchmaschinen.htm> [Download 13.02.02].
- V-business (ohne Jahr): Zukunft der Suchmaschinen. URL: <http://www.suchmaschinen-marketing.de/zukunftsuchmaschinen.htm> [Download 13.02.02].
- Vowe, Gerhard & Martin Emmer (2001): «Digital Divide»: Ein Mythos auf dem Prüfstand. In: *Politik im Netz (PIN)* URL: <http://www.politik-im-netz.com> [Download 13.02.02].
- Weischenberg, Siegfried (1995): *Journalistik. Theorie und Praxis aktueller Medienkommunikation*. Opladen.

#### Anmerkungen

- <sup>1</sup> Hack (1998: 191) vergleicht die Suchmaschinen mit Nachrichtenagenturen und die Softwareagenten mit den Korrespondenten.
- <sup>2</sup> Das Geschäft mit den Meta-Informationen (Suchmaschinen) ist vielfach attraktiver als das Geschäft mit den Informationen selbst (vgl. Kuhlen 1999: 156).
- <sup>3</sup> Im Internet findet man eine Vielzahl von Informationsquellen zum Thema Suchmaschinen (zur Technik, zur Funktionsweise, zur Entwicklung, zu den Perspektiven...). Neben den im Text explizit genannten Quellenangaben hier noch drei weitere Hinweise auf ausführliche und informative Netzquellen:
  1. Searchenginewatch (ohne Jahr) URL: <http://www.searchenginewatch.com>
  2. Bekavac, Bernard (2001): Tutorial zur Suche im WWW/Internet, Methoden und Verfahren von Suchdiensten im

WWW/Internet. URL: <http://www.inf-wiss.uni-konstanz.de/suche/3>. Die Suchfibel (ohne Jahr): URL: <http://www.suchfibel.de>

- <sup>4</sup> Potempa et al. (1998: 74 ff.) unterscheiden zwischen Gateways, globalen Suchmaschinen, Meta-Suchern und speziellen Suchmaschinen. Alle vier Typen werden wiederum in mehrere Subtypen unterteilt. Zur Abgrenzung werden implizit inhaltliche, technische und strukturelle Kriterien herangezogen. Aber auch bei dieser Einteilung sind die Übergänge fließend. Für die zukünftige Entwicklung relevant ist sicherlich vor allem die Unterscheidung von globalen und spezialisierten Suchmaschinen. Die Bedeutung spezialisierter Anbieter wird sicherlich mit zunehmender Größe des Netzes und mit sinkender Ausschöpfung durch die universellen Anbieter wachsen (vgl. Gartz 2000: 46). Je stärker aber solche Spezialisierungen sich verbreiten, desto dringender wird wiederum die Notwendigkeit von Meta-Informationen, in diesem Fall die Suchmaschine für die geeignete Suchmaschine.
- <sup>5</sup> Während die Zuordnungskategorien immer durch Menschen festgelegt werden, wird die anschließende Einordnung einzelner Dokumente in diese Kategorien nicht unbedingt durch Redakteure vorgenommen. Man lässt hier – ähnlich wie bei der Indizierung durch Robots – auch eine automatisierte Zuordnung anhand der vorgegebenen Begriffe durchführen (vgl. Hartmann, Näf & Schäuble 2000: 112 ff.).
- <sup>6</sup> Zur Definition und zu den Anforderungen, die an echte Meta-suchmaschinen zu stellen sind, vgl. Kuhlen (1999: 244) sowie Gartz (2000: 51).
- <sup>7</sup> Dabei haben die verschiedenen Suchmaschinen unterschiedliche Strategien, insbesondere hinsichtlich der Indizierungstiefe und -breite (vgl. Philippus 1997: 68). Große Unterschiede sind auch bezüglich der vorbereitenden Schritte festzustellen: Dazu gehören Sprachidentifikation, Buchstabenumwandlungen, Stoppwortelimination, Wortnormalisierung etc., alles Routinen, die sich später auf die Qualität der Retrievals auswirken (vgl. Hartmann, Näf & Schäuble 2000: 51 ff.).
- <sup>8</sup> Zu den Aspekten des Relevanzbegriffs – subjektive, objektive und geschätzte Relevanz – die im Zusammenhang mit Suchmaschinen unterschieden werden, vgl. Hartmann, Näf & Schäuble (2000: 34 ff.).
- <sup>9</sup> Zum einen kann man die Anmeldung selbst durchführen oder von einem professionellen Registrierungsdienst gegen Gebühr vornehmen lassen, zum anderen gibt es aber auch Softwarelösungen, die die Anmeldung bei mehreren Suchmaschinen automatisiert vornehmen (Gartz 2000: 155 ff.). In jedem Fall ist die Anmeldung für den Content-Anbieter aber mit Kosten verbunden, zeitlichen und/oder finanziellen. Als Ausgangspunkt für die Suche spielen neben den Anmeldungen vor allem bereits bestehende Adressenbestände (z.B. die Adressenlisten eines Kataloges) eine zentrale Rolle (Potempa et al. 1998: 200 f.). Durch beide Aspekte wird die Auffindbarkeit von Web-Seiten finanzstarker Anbieter systematisch bevorzugt.
- <sup>10</sup> Dennoch gibt es einige Unterschiede: Manche Suchdienste beschränken die Suchtiefe bzw. die Seitenzahl pro Web-Seite. Das hat zur Folge, dass in der Verzeichnishierarchie tieferliegende Informationen nicht gefunden werden. Trotz der globalen Struktur des WWW werden offenbar Web-Seiten auf US-Servern von vielen Suchmaschinen bevorzugt aufgesucht, gefunden und indiziert (Gartz 2000: 48). Des Weiteren unterscheiden sich die Maschinen darin, welche zusätzlichen Ressourcen neben HTML-Dokumenten von den Suchrobotern berücksichtigt werden (vgl. Potempa et al. 1998: 102 ff.).
- <sup>11</sup> Für die Nutzer stehen solche Filter bereits zur Verfügung und werden eingesetzt. Sie können verwendet werden, um Kinder den Zugang zu Gewalt, Pornographie oder Drogenthemen zu verwehren. Sie können auch Büroangestellte während der

Arbeitszeit von Urlaubs-, Sport- und Shoppingangeboten fernhalten (vgl. Kuhlen 1999: 267). Die gleichen Techniken lassen sich natürlich auch in Suchmaschinen implementieren, und es gibt keine Garantie dafür, dass die Filter nur auf Sex, Drogen und Gewaltverherrlichung beschränkt bleiben.

<sup>12</sup> Die Suchmaschinen unterscheiden sich allerdings darin, ob sie das gesamte Dokument durchsuchen oder nur den Anfang. Nicht einheitlich ist auch, welche Worte sie überhaupt indizieren (vgl. Philippus 1997: 70).

<sup>13</sup> Dies ist anschaulich dargestellt bei Hartmann, Näf & Schäuble (2000: 67 f.).

<sup>14</sup> Bei einer Suchanfrage mit mehreren Begriffen ist es natürlich vor allem wichtig, wie viele der eingegebenen Suchbegriffe im Dokument enthalten sind und wie ausgefallen (selten) diese jeweils sind. Darüber hinaus ist der Abstand zwischen den Suchbegriffen im Text für die Platzierung der Seite von Bedeutung (vgl. Hartmann, Näf & Schäuble 2000: 39 f.).

<sup>15</sup> Allerdings sind leistungsfähige Programme in der Lage, Synonyme zu identifizieren und bei der Indizierung zu berücksichtigen, sodass dieses Problem reduziert werden kann (vgl. Hack 1998: 197). Um die zahlreichen Schwierigkeiten, die bei der Schlüsselwortindizierung auftreten, in den Griff zu bekommen, versucht man stattdessen eine konzeptbasierte Indizierung durchzuführen, d.h. durch Wortfelder, Wortabstandsrechnungen etc. den Inhalt eines Dokuments zu erfassen (vgl. Philippus 1997: 71; Potempa et al. 1998: 120 f.).

<sup>16</sup> Dieses als kollaboratives Filtern bezeichnete Verfahren wird auch verwendet, um den Usern bei Onlinekäufen gezielte Produktvorschläge zu machen. Das Verhalten anderer Benutzer wird damit zur Basis für die aktuelle Selektionsentscheidung der Maschine (vgl. Kuhlen 1999: 256 ff.).

<sup>17</sup> Aus gesellschaftlicher Perspektive ist es durchaus positiv einzuschätzen, dass durch die verschiedenen Suchmaschinen unterschiedliche Retrievals produziert werden und somit eine gewisse Vielfalt gewährleistet wird. Allerdings haben die meisten User offenbar einen Favoriten unter den Suchmaschinen, den sie immer wieder nutzen. Von daher bekommen sie dennoch nur eine Sicht auf die «WWW-Welt» (vgl. Kuhlen 1999: 247).

<sup>18</sup> Auch in den einschlägigen Handbüchern finden sich Tipps und Hinweise, worauf zu achten ist, wenn man bei der Indizierung durch die Suchmaschinen die Relevanz der eigenen Seite erhöhen möchte (vgl. Potempa et al. 1998: 126 ff.; Gartz 2000: 147 ff.). Auf die damit verbundenen Probleme wird aber nicht oder nur am Rande eingegangen.

<sup>19</sup> Aus der Userperspektive haben solche Algorithmen auf den ersten Blick durchaus positive Effekte auf die Qualität von Suchergebnissen. Andererseits wird die Offenheit des Netzes damit aber in Frage gestellt. Die etablierten Filter werden auf Umwegen reaktiviert.