

2 Untersuchungsgegenstand Suchmaschinen

Suchmaschinen gehören zu der Art von Webdiensten, deren Sinn und Nutzen sich in der täglichen Praxis erschließt, ohne dass man sich vertieft mit Begrifflichkeiten oder ihrer technischen Arbeitsweise beschäftigt. Dennoch erscheint ein gewisses Systemverständnis für den weiteren Verlauf der Arbeit entscheidend. Für die grundrechtliche Analyse muss dieses Wissen gewissermaßen als Subtext vorausgesetzt werden.

Im Folgenden sollen daher Begriffsfragen geklärt werden. Außerdem soll die technische Wirkweise von Suchmaschinen, insbesondere das Wesen des Algorithmus als zentraler Bestandteil, dargestellt werden.

2.1 Vorfragen

2.1.1 Der Begriff der Suchmaschine

In Ermangelung einer verbindlichen Terminologie ist nicht abschließend klar, welche Arten von Onlinediensten sich potentiell unter den Begriff der Suchmaschine fassen lassen. Für diese Arbeit jedenfalls gilt das heute im allgemeinen Sprachgebrauch²¹ herrschende Verständnis von Suchmaschinen, welches so im Ergebnis auch von der Rechtswissenschaft geteilt wird.²² Eine Suchmaschine ist demnach akkurat ausgedrückt:

„ein Computersystem, das verteilte Inhalte aus dem World Wide Web mittels Crawling erfasst und über eine Benutzerschnittstelle durchschaubar macht, wobei die Ergebnisse in einer nach systemseitig angenommener Relevanz geordneten Darstellung aufgeführt werden.“²³

Üblich ist grundsätzlich die Eingabe von Stichworten in ein Suchformular. Nach erfolgter Eingabe wird als neue Seite die Trefferliste aufgebaut. Diese zeigt mit kurzen Erläuterungen bzw. einer Inhaltsvorschau versehene Hyperlinks²⁴ an. Folgt man einem der Hyperlinks, gelangt man – jedenfalls im Idealfall – auf den gesuchten Inhalt. Auch

²¹ „Such|ma|sch|ne, die: auf einen bestimmten Namen lautendes Programm im Internet, das mithilfe umfangreicher, aus Internetadressen bestehender Datenbanken die gezielte Suche nach Informationen im Internet ermöglicht“, aus Scholze-Stubenrecht, Duden – Deutsches Universalwörterbuch, 72011; vgl. auch Lewandowski, Web Information Retrieval, 2005, S. 25.

²² Vgl. statt vieler Eberle, Mainzer Rechtshandbuch der neuen Medien, 2003, S. 14; Kühling/Gauß, ZUM 2007, 881, 882; Machill/Beiler/Zenker, in: Machill (Hrsg.), Die Macht der Suchmaschinen, 2007, S. 8; Brunn, Cache me if you can, 2013, S. 19 f.

²³ Lewandowski, Suchmaschinen verstehen, 2015, S. 29.

²⁴ Vgl. weiterführend zur Idee des Webs als Netzwerk und der technischen Struktur der Hyperlinks Witten/Numerico/Gori, Web Dragons, 2006, S. 61 ff.

wenn die klassische Eingabe eines Stichwortes der Normalfall sein dürfte, so sind mittlerweile auch Spracheingaben oder beispielsweise Standortdaten via GPS als Suchanfragen denkbar.²⁵ Als typisierter Normalfall soll aber im Folgenden nur die klassische Stichwortsuche Betrachtungsgegenstand sein.²⁶

Ebenso wird im Kontext dieser Arbeit, vorbehaltlich entsprechender Kennzeichnung, als Suchmaschine lediglich die automatisierte, indexbasierte Suchmaschine bezeichnet. Darunter fasst man nur jene Angebote, welche einen eigenen Index, also eine eigene Datenbank, zur Verarbeitung von Suchanfragen verwenden.²⁷

Demgegenüber stehen Meta-Suchmaschinen. Diese nutzen die Indices bestehender Drittangebote und leiten Sucheingaben an diese weiter, um aus den aggregierten Ergebnissen der angesteuerten Dritt-Suchmaschinen eine einheitliche „eigene“ Trefferliste zu erzeugen.²⁸

Im Fokus der folgenden Betrachtungen steht die Frage nach der Meinungsbildungsrelevanz von Suchmaschinen. Entscheidend dafür ist, ob und in welchem Umfang Suchmaschinen durch eine eigene Wertung und Selektionsleistung eine aktive Rolle im Prozess der Informationsvermittlung spielen. Meta-Suchmaschinen übersetzen Suchanfragen in verschiedener Weise in für die angesteuerten Suchmaschinen lesbare Eingaben und sind damit natürlich in gewisser Weise selektiv.²⁹ Eine echte originäre Eigenleistung bezüglich der Wertung und Selektion von Inhalten setzt aber eben einen eigenen Index, einen eigenen Ansatz in der Informationsstrukturierung, voraus.

Im Übrigen sind Meta-Suchmaschinen in ihrer Konzeption denklogisch von den Leistungen anderer Suchmaschinen abhängig. Dementsprechend sind sie als potentiell

²⁵ Vgl. <http://www.google.com/intl/de/insidesearch/features/voicesearch/> (Zugriff am 16.02.2016).

²⁶ Schulz/Held/Laudien, Suchmaschinen als Gatekeeper in der öffentlichen Kommunikation, 2005, S. 13 f. und Paal, Suchmaschinen, Marktmacht und Meinungsbildung, 2012, S. 19 verweisen noch auf Webkataloge und -verzeichnisse als Unterfall von Suchmaschinen; diese dürften mittlerweile im Sprachgebrauch nicht mehr als „Suchmaschinen“ verstanden werden, vgl. Griesbaum/Bekavac/Rittberger, in: Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen, 2009, S. 18; so im Ergebnis auch das LG Mannheim schon im Jahre 1998, LG Mannheim, MMR 1998, 217, 217. Jedenfalls haben sie kaum noch Bedeutung, so Hürlimann, Suchmaschinenhaftung, 2012, S. 7.

²⁷ Zum Begriff des Index siehe Kapitel 2.2.2.

²⁸ Türker, The optimal design of a search engine from an agency theory perspective, 2004, S. 10

²⁹ Türker, The optimal design of a search engine from an agency theory perspective, 2004, S. 10.

wirkmächtige Akteure von geringerer eigenständiger Bedeutung als Suchmaschinen mit eigenem Index.³⁰

2.1.2 Suchmaschinen als Blackbox-Systeme

Suchmaschinen sind im Grunde als Blackbox-Systeme konzipiert. Ihr spezifischer Aufbau hängt stark vom jeweiligen Anbieter ab und ist nicht publik. Schon aus wirtschaftlichem Eigeninteresse der Betreiber ist auch nicht damit zu rechnen, dass er jemals öffentlich wird.

Dementsprechend sind kaum Informationen aus erster Hand verfügbar.³¹ Als Konsequenz kann die folgende Arbeit keine Erläuterungen zu technischen Details oder rechtlichen Implikationen einzelner Angebote geben. Sie widmet sich grundsätzlichen Fragen der Gestaltung von Suchmaschinen und des Wesens algorithmischer Prozesse. Durch die Wissenschaft und letztendlich Reverse-Engineering-Ansätze³² können diese als gesichert betrachtet werden. Sie bilden die Grundlage der rechtlichen Überlegungen.

Im Übrigen handelt es sich bei Suchmaschinentechnologien der Sache nach um Information-Retrieval-Modelle. Diese widmen sich der Strukturierung, Organisation und Gewinnung von Informationen.³³ Dieses Forschungsfeld ist an sich weit älter als das Web.³⁴ Suchmaschinen stellen nun den praktischen Anwendungsfall dieser Modelle auf große Textsammlungen dar.³⁵ Insofern können die Erkenntnisse dieser Disziplin herangezogen werden.

³⁰ So bereits *Elixmann*, Datenschutz und Suchmaschinen, 2012, S. 45; grundlegend zur „Zweitrangigkeit“ von Metasuchmaschinen *Lewandowski*, Web Information Retrieval, 2005, S. 25 f.; *ders.*, Suchmaschinen verstehen, 2015, S. 22, der dort sogar den Nutzen von Metasuche allgemein anzweifelt.

³¹ So bereits *Brunn*, Cache me if you can, 2013, S. 22 dort Fn. 26; eine – wenn auch generisch gehaltene – Ausnahme bildet der Aufsatz der Google-Gründer selbst, vgl. *Brin/Page*, Computer Networks and ISDN Systems 1998, 107 ff.

³² Hier ist vor allem an die Berufsparte der SEO (Search Engine Optimization) zu denken, vgl. *Winston*, Dalhousie Journal of Interdisciplinary Management 2011, 1, 3.

³³ Vgl. *Croft/Metzler/Strohman*, Search engines, 2010, S. 1.

³⁴ Vgl. *Croft/Metzler/Strohman*, Search engines, 2010, S. 1 ff.

³⁵ *Croft/Metzler/Strohman*, Search engines, 2010, S. 6.

2.1.3 Suchmaschinen und Google

Als unstreitiger Marktführer unter den Suchmaschinen nimmt die Suchmaschine Google³⁶ eine hervorgehobene Rolle ein.³⁷ Aufgrund ihrer Marktdominanz, ihrer stetig wachsenden Verbreitung – man denke nur an das Betriebssystem Android – und des nahezu synonymen Verständnisses von „googeln“ als „im Internet suchen“³⁸ lassen sich Rechtsfragen unter dem Topos „Machtfaktor Suchmaschine“ exemplarisch gut an Google erörtern.

Im Schrifttum stand Google dementsprechend bisher auch klar im Blickpunkt.³⁹ Die folgenden Ausführungen betrachten Suchmaschinen als Beispiel für den Einfluss von Algorithmen auf die Meinungsbildung im Allgemeinen. Allerdings wird aus den genannten Gründen auch hier Google vorrangig als Realbeispiel herangezogen.

³⁶ Die Suchmaschine der Google Inc., bald Alphabet Inc.; der Begriff „Google“ wird teilweise auch als Begriff für den gesamten Konzern verwendet.

³⁷ Vgl. nur *Kilian/Heussen*, Computerrechts-Handbuch, 2013, 3. Abschnitt, Suchmaschinen Rn. 20 m. w. N.; der Marktanteil weltweit liegt bei ca. 68 % (Desktopnutzung) bzw. 92 % (Mobilnutzung), <http://de.statista.com/statistik/daten/studie/222849/umfrage/marktanteile-der-suchmaschinen-weltweit/> (Zugriff am 16.02.2016); in Deutschland sogar bei ca. 91 % (Desktopnutzung) und sogar 97 % (Mobilnutzung), <http://de.statista.com/statistik/daten/studie/301012/umfrage/marktanteile-der-suchmaschinen-und-marktanteile-mobile-suche/> (Zugriff am 16.02.2016).

³⁸ „*googeln [...] mit Google im Internet suchen, recherchieren: ich goog[e]le mal schnell; er hatte ihren Namen, nach Informationen gegoogelt*“, aus *Scholze-Stubenrecht*, Duden – Deutsches Universalwörterbuch, 2011. In der vorangegangenen Ausgabe des Dudens war noch allgemein von „Suchen im Internet“ die Rede; vgl. zum markenrechtlichen Hintergrund eines allgemeinen Verbes „googeln“ und den Befürchtungen der Google Inc. *Parbel*, Google will richtig „googeln“ lassen, Heise Online am 15.08.2006, abrufbar unter <http://www.heise.de/newsticker/meldung/Google-will-richtig-googeln-lassen-152201.html> (Zugriff am 16.02.2016).

³⁹ Vgl. *Kühling/Gauß*, MMR 2007, 751, 751 ff.; *Kühling/Gauß*, ZUM 2007, 881, 881 ff.; *Hopf*, Der Missbrauch einer marktbeherrschenden Stellung von Internetsuchmaschinen, dargestellt am Beispiel von Google, 2014; *Danckert/Mayer*, MMR 2010, 219, 219 ff.; *Ott*, Die vorherrschende Meinungsmacht von Google – Eine Replik zu Danckert/Mayer, MMR 2010, 219 ff., MMR-Aktuell 2010, 301459; *Körber*, WRP 2012, 761, 761 ff.; für englischsprachige Literatur exemplarisch *Lopez-Tarruella* (Hrsg.), *Google and the Law*, 2012; *Ghatnekar*, 33 *Loyola of Los Angeles Entertainment Law Review* 171 (2013), 171 ff.

2.2 Der Aufbau im Detail

Um die Arbeitsweise einer Suchmaschine modelhaft nachvollziehen zu können, bietet sich eine konzeptionelle Dreiteilung an.⁴⁰ Es soll zwischen der Inhaltserfassung (Crawling), dem Auswertungs- und Strukturierungsprozess in einem Index definierter Daten (Indexierung) und schließlich der Abfrage als Reaktion auf eine Sucheingabe des Nutzers differenziert werden (Abfrage). Letztere führt zu einer nach Relevanzkriterien erstellten Trefferliste.

2.2.1 Crawling

Zunächst muss eine Suchmaschine den vorhandenen Inhalt des Webs erfassen. Dazu wandert ein Programm, der Crawler,⁴¹ automatisch über Verlinkungen von URL⁴² zu URL und lädt die Bestandteile der besuchten Seite in einen Speicher. Bei seiner automatischen Wanderung nutzt er die Netzstruktur des Webs. Auf jeder besuchten URL werden ausgehende Links analysiert, in eine Liste aufgenommen und dann entsprechend abgearbeitet.⁴³ Ausgehend vom bereits erfassten Inhalt ist es zentrale Aufgabe des Crawlers, neue, geänderte, gelöschte oder verschobene Dokumente zu ermitteln.⁴⁴ Der Ursprung des Crawlings war die simple Frage nach der Größe des Webs. Der World Wide Wanderer, als erster Crawler überhaupt, sollte über das Abwandern der gesamten Netzstruktur genau dies herausfinden.⁴⁵

Für Suchmaschinen ist eine zufällige, von einer beliebigen URL beginnende Wanderung durchs Web wenig effizient. Schon aus Kapazitätsgründen muss eine Entscheidung erfolgen, welche Seiten priorisiert angesteuert werden (Page Selection). Außerdem ist zu entscheiden, welche URLs in welchen Intervallen erneut abgerufen werden,

⁴⁰ So auch zum Folgenden *Griesbaum/Bekavac/Rittberger*, in: Lewandowski (Hrsg.), *Handbuch Internet-Suchmaschinen*, 2009, S. 28; ebenso statt vieler *Machill/Beiler/Zenker*, in: Machill (Hrsg.), *Die Macht der Suchmaschinen*, 2007, S. 8; *Elixmann*, *Datenschutz und Suchmaschinen*, 2012, S. 41 ff.; für einen weiter aufgeschlüsselten Aufbau vgl. *Lewandowski*, *Web Information Retrieval*, 2005, S. 26 ff.

⁴¹ Auch Spider, Robot oder Searchbot genannt, im Rahmen dieser Arbeit aber einheitlich als Crawler bezeichnet.

⁴² Abkürzung für Uniform Resource Locator.

⁴³ *Witten/Numerico/Gori*, *Web Dragons*, 2006, S. 70.

⁴⁴ *Lewandowski*, *Web Information Retrieval*, 2005, S. 49.

⁴⁵ Von *Mathew Gray* am MIT entwickeltes Skript, um die Größe des Webs zu messen, siehe <http://web.archive.org/web/20030512083018/http://ksi.cpsc.ualgary.ca/archives/WWW-TALK/www-talk-1993q2.messages/706.html> (Zugriff am 16.02.2016).

um die Aktualität der Erfassung zu gewährleisten (Page Refresh).⁴⁶ Die Abdeckung aller Webinhalte, der Aktualität der Erfassung und das Ausmaß der Übereinstimmung der erfassten Inhalte mit dem im jeweiligen Moment tatsächlich vorhandenen Webinhalt prägen die Qualität einer Suchmaschine.⁴⁷

Ein perfektes Ergebnis ist diesbezüglich nicht möglich. Es bedarf daher vonseiten der Anbieter konkreter Strategieentscheidungen bei der Gestaltung ihres Crawling-Programms.⁴⁸ So bewarb beispielsweise Google im Rahmen eines Updates ihrer Suchmaschine explizit eine Verbesserung der Aktualität (Freshness).⁴⁹

Ein Problem der Inhaltserfassung über Crawler ist die Interaktion mit Webservern. Grundsätzlich kommuniziert der Crawler über das Hypertext-Transfer-Protokoll (http). Er zeichnet so den Weg eines menschlichen Surfers nach. Nicht über die gewöhnliche Link-Struktur auffindbarer Webinhalt, wie beispielsweise dynamische Webseiten, die als Antwort auf eine spezifische Eingabe erschaffen werden, ist demnach nicht katalogisierbar. Selbiges gilt für passwortgeschützte Inhalte wie zum Beispiel Datenbanken.⁵⁰

Daneben bietet das Robot-Exclusion-Standard-Protokoll die Möglichkeit, das Crawlen einer URL zu verhindern. Mit Hilfe einer robots.txt-Datei auf dem jeweiligen Webserver kann der Serverbetreiber festlegen, welche Bereiche auf seinem Server vom Crawler nicht erfasst werden. Die diesbezügliche Festlegung bedingt keinen technischen Ausschluss, wird aber von den Suchmaschinenanbietern als Marktstandard akzeptiert.⁵¹

2.2.2 Indexierung

Der Index stellt jene Datenbank dar, welche bei der konkreten Suchanfrage nach Treffern analysiert wird. Es findet also beim Suchvorgang, wie schon logisch aus der Exis-

⁴⁶ Vgl. *Türker*, The optimal design of a search engine from an agency theory perspective, 2004, S. 15.

⁴⁷ Coverage, Freshness und Coherence nach *Witten/Numerico/Gori*, Web Dragons, 2006, S. 83 f.

⁴⁸ Vgl. *Witten/Numerico/Gori*, Web Dragons, 2006, S. 70.

⁴⁹ Vgl. <http://googleblog.blogspot.de/2010/06/our-new-search-index-caffeine.html> (Zugriff am 16.02.2016).

⁵⁰ *Witten/Numerico/Gori*, Web Dragons, 2006, S. 71; dieser als „Deep Web“ bezeichnete Teil des Webs soll sogar mehr Dokumente enthalten als der von Suchmaschinen erfassbare Teil, vgl. *Lewandowski*, Suchmaschinen verstehen, 2015, S. 235.

⁵¹ *Witten/Numerico/Gori*, Web Dragons, 2006, S. 71; *Lewandowski*, Suchmaschinen verstehen, 2015, S. 41.

tenz des Crawling-Vorgangs folgt, grundsätzlich keine Echtzeit-Abfrage des Netzes statt.⁵²

Der Aufbau des Suchindex lässt sich grundlegend in die Teilbereiche der Datennormalisierung, der Datenanalyse und den Aufbau durchsuchbarer Dateistrukturen untergliedern.⁵³ Die ersten beiden Schritte sind grundlegende Voraussetzung für Letzteren.

Im Rahmen der Normalisierung werden die gefundenen Webinhalte in ein einheitliches Format versetzt und auf ihre wesentliche Struktur reduziert. Es gilt die Fülle der Inhalte in ein automatisiert bearbeitbares Format zu wandeln.

Die Datenanalyse schlüsselt den gefundenen Inhalt nach Merkmalen auf. Grundlegendes Element dabei sind die Wörter eines Dokuments, für welche ein textbasierter, invertierter Index angelegt wird. Dabei handelt es sich in seiner einfachsten Form um eine Liste mit Wörtern und der Angabe, auf welcher Seite und an welcher Position dort das jeweilige Wort zu finden ist. Darüber hinaus finden in der Praxis freilich weit komplexere Verarbeitungsvorgänge statt. Beispielsweise lassen sich Wörter auf Wortgruppen bzw. Stammformen reduzieren. Auch können der semantische Kontext eines Wortes, die Platzierung (Überschrift) oder eine besondere Schriftart (Fett, Kursiv etc.) analysiert werden.⁵⁴

Daneben werden alle erdenklichen sonstigen Informationen für weitere Indices angelegt. Als bekannte Beispiele sind das Erstelldatum der Website, Dokumentenlängen, Server-Informationen, Meta-Tag-Informationen und Informationen aus Netzstruktur des Webs wie die Anzahl der ein- und ausgehenden Links zu benennen.⁵⁵

Grundsätzlich soll der Index als strukturierte Datenbank eine Kommunikation mit in ein passendes Dateiformat übersetzten Suchanfragen ermöglichen.⁵⁶ Der Index muss so aufgebaut sein, dass ein Abgleich mit Merkmalen möglich ist, die aus der Suchanfrage entnommen werden können. Dementsprechend hängt die genaue Gestaltung des Index wechselseitig vom für die Relevanzbestimmung verantwortlichen Ranking-

⁵² Lewandowski, Suchmaschinen verstehen, 2015, S. 49.

⁵³ So Hürlimann, Suchmaschinenhaftung, 2012, S. 16; die Unterteilung ist dabei nicht zwingend; feingliedriger zum Beispiel Brunn, Cache me if you can, 2013, S. 28 ff.

⁵⁴ Vgl. Türker, The optimal design of a search engine from an agency theory perspective, 2004, S. 16 f.; Witten/Numerico/Gori, Web Dragons, 2006, S. 104 ff.

⁵⁵ Vgl. Griesbaum/Bekavac/Rittberger, in: Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen, 2009, S. 32 f.; Türker, The optimal design of a search engine from an agency theory perspective, 2004, S. 16.

⁵⁶ Lewandowski, Suchmaschinen verstehen, 2015, S. 48.

Algorithmus und der Abfrage ab.⁵⁷ Bei einem Ranking-Algorithmus, der Relevanz einzig anhand der Häufigkeit eines Wortes in einem Dokument bestimmt, wäre die Textanalyse und damit ein einfacher Textindex ausreichend.

2.2.3 Abfrage

Während die Inhaltserfassung und Aufbereitung die Grundlagen für den Erfolg einer Suchmaschine legen, ist es die Abfrage, welche eine Nutzerinteraktionskomponente in Form der Suchanfrage enthält und zur konkreten Trefferliste führt. Im Ergebnis ist damit die Abfrage für die Resonanz der Nutzer einer Suchmaschine entscheidend.

2.2.3.1 Das Abfragemodul

Ein Abfragemodul (Query Module) übersetzt zunächst die Suchanfrage des Nutzers (in der Regel einige Wörter) in eine für die Index-Datenbank verständliche Form. Dabei wird im Prinzip die gleiche Art der Textverarbeitung angewandt, die auch der Indexerstellung zugrunde liegt.⁵⁸ Ziel ist es, den tatsächlichen Aussagegehalt einer Suchanfrage, mithin also die eigentliche Fragestellung zu identifizieren.

Die bestehende Suchmaschinen-Architektur ist dabei auf kurze Schlüsselwörter fokussiert. Längere Suchanfragen und Satzstrukturen werden noch ungenau verarbeitet.⁵⁹ Damit korrespondiert allerdings auch eine im Durchschnitt nur wenig ausgeprägte Bereitschaft der Nutzer, für die Formulierung ihrer Suchanfragen besondere Mühen aufzuwenden. Die Nutzer verlassen sich im Ergebnis auf die internen Mechanismen des Systems Suchmaschine.⁶⁰ Zur schrittweisen Lösung dieses Problems wird hierzu beispielsweise auf das Nutzerverhalten zurückgegriffen. Suchmaschinen analysieren, welche Treffer für eine bestimmte Suchanfrage tatsächlich angeklickt wurden und sind somit immer besser in der Lage, den Sinngehalt von Suchanfragen zu verstehen.⁶¹ Außerdem geben Suchmaschinen ihren Nutzern mittlerweile ad hoc bei der Eingabe von

⁵⁷ Griesbaum/Bekavac/Rittberger, in: Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen, 2009, S. 32; Türker, The optimal design of a search engine from an agency theory perspective, 2004, S. 16; vgl. auch Brin/Page, Computer Networks and ISDN Systems 1998, 107, 110.

⁵⁸ Croft/Metzler/Strohman, Search engines, 2010, S. 24.

⁵⁹ Croft/Metzler/Strohman, Search engines, 2010, S. 23 f.

⁶⁰ Lewandowski, Suchmaschinen verstehen, 2015, S. 74 f.

⁶¹ Türker, The optimal design of a search engine from an agency theory perspective, 2004, S. 23; auf diese Weise verbessert sich gewissermaßen die Qualität einer Suchmaschine durch eine entsprechend hohe Nutzerzahl.

Suchbegriffen Vorschläge an die Hand und komplettieren als unvollständig analysierte Eingaben um eigene Ergänzungen (Autocomplete).⁶²

Die so aufbereitete Suchanfrage wird nun mit dem Index abgeglichen und schlussendlich nach bestimmten systemseitigen Gewichtungskriterien eine Trefferliste erstellt. Für begleitende Fragen, wie die grafische Darstellung (zum Beispiel Hervorhebungen in der Schriftart, Erstellung von kleinen Vorschautexten, sogenannten Snippets) ist eine separate Result-Output-Komponente verantwortlich.⁶³

2.2.3.2 Die Gewichtung

Die Gewichtung oder auch Ranking, also die Frage der Rangfolge der angezeigten Treffer, wird von einem sogenannten Ranking-Modul übernommen.⁶⁴ Anhand komplexer Algorithmen⁶⁵ versucht sich die Suchmaschine an einer Gewichtung der Relevanz von Webseiten mit Blick auf die zuvor identifizierte Fragestellung. Das genaue Wie der Gewichtung stellt dabei das „Herz“ der Suchmaschine dar. Als zentrale Elemente lassen sich der Inhalt des Dokuments, seine logische Struktur, externe Faktoren und das Nutzerverhalten identifizieren.⁶⁶

Im Hinblick auf die Merkmale der einzelnen Dokumente selbst liegt dem Ranking prinzipiell eine klassische Textanalyse zugrunde. Der Index wird beispielsweise auf Dokumente untersucht, die alle Suchwörter enthalten. Weiterhin können Merkmale wie Groß- und Kleinschreibung, die Nennung in der URL selbst und die Position der Suchwörter im Text in die Bewertung einfließen.⁶⁷ Als Ergebnis wird ein Score-Wert für einzelne Dokumente berechnet, aus dem sich eine grundsätzliche Reihenfolge ergibt.

Nun ist die bloße Textanalyse kein hinreichendes Mittel, um in der Überfülle des Web brauchbare Treffer zu generieren.⁶⁸ Es bedarf der Heranziehung externer Faktoren – der Dokumentenumgebung –, die mit den Werten der reinen Textanalyse verrechnet werden. Als prominentestes Beispiel für die Berücksichtigung des Dokumentenumfel-

⁶² Lewandowski, Suchmaschinen verstehen, 2015, S. 67.

⁶³ Croft/Metzler/Strohman, Search engines, 2010, S. 25.

⁶⁴ Türker, The optimal design of a search engine from an agency theory perspective, 2004, S. 18.

⁶⁵ Zum Begriff des Algorithmus siehe unten Kapitel 4.3.

⁶⁶ Türker, The optimal design of a search engine from an agency theory perspective, 2004, S. 18 ff.

⁶⁷ Vgl. Brin/Page, Computer Networks and ISDN Systems 1998, 107, 113; Lewandowski, Suchmaschinen verstehen, 2015, S. 92.

⁶⁸ Griesbaum/Bekavac/Rittberger, in: Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen, 2009, S. 34.

des gilt wohl Page Rank. *Brin* und *Page* entwickelten einen Algorithmus, der die Linkstruktur des Webs selbst zur Gewichtung von Webseiten heranziehen sollte. Im Grundsatz erfolgt so eine Übernahme des wissenschaftlichen Zitationsgedankens in das Feld der Suchmaschinentheologie. Webseiten, auf die besonders häufig verwiesen wird – also solche, die häufig zitiert werden –, sollen bedeutsamer sein als andere. Diese Methodik wird als objektives Verfahren präsentiert, welches sich besonders gut mit der subjektiven Vorstellung des Einzelnen von Relevanz deckt.⁶⁹ Dabei ist nicht nur die Häufigkeit der eingehenden Links entscheidend, sondern auch der Page Rank der verlinkenden Seite.

Prinzipiell errechnet sich der Page Rank einer Seite aus der Addition der eingehenden Links, welche jeweils mit Punktwerten versehen sind. Der Punktwert eines jeden Links variiert dabei. Er bestimmt sich aus einer Division des Page Ranks der verlinkenden Seite durch die Anzahl der ausgehenden Links.⁷⁰ Hinzu kommt ein Zufallsfaktor (*d*) (auch Dämpfungsfaktor genannt) zwischen 0 und 1, welcher der Wahrscheinlichkeit entspricht, dass ein zufälliger Surfer auf eine zufällige andere Seite wechselt und nicht weiter der Linkstruktur folgt.⁷¹ Dieser Wert lässt sich einzelnen Seiten, aber auch Seitengruppen hinzufügen und erlaubt eine starke Justierung des Page Ranks, losgelöst vom eigentlichen Zitationsgedanken.⁷² Weiterhin ist für das Page-Rank-Verfahren zu beachten, dass der Page Rank einer Seite fest zugeordnet und in der Grundkonzeption unabhängig von bestimmten Suchanfragen ist.⁷³ Page Rank ermittelt also für sich genommen nur die bestimmte Qualität einer Seite, nicht die Relevanz für eine bestimmte Suchanfrage.⁷⁴

Darüber hinaus ordnet Google die Information aus dem Ankertext⁷⁵ eines Links nicht mehr der Seite zu, auf der sich besagter Link befindet, sondern der Seite, auf die er

⁶⁹ *Brin/Page*, Computer Networks and ISDN Systems 1998, 107, 110; eine Annahme, die mit Blick auf den Zufallsfaktor (*d*) in Frage gestellt werden kann, siehe dazu Fn. 72. Kritisch auch *Griesbaum/Bekavac/Rittberger*, in: Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen, 2009, S. 37.

⁷⁰ Grundsätzlich dazu *Brin/Page*, Computer Networks and ISDN Systems 1998, 107, 109 ff.; weiterführend *Witten/Numerico/Gori*, Web Dragons, 2006, S. 113 ff.

⁷¹ *Brin/Page*, Computer Networks and ISDN Systems 1998, 107, 110.

⁷² *Witten/Numerico/Gori*, Web Dragons, 2006, S. 117; *Brin/Page*, Computer Networks and ISDN Systems 1998, 107, 110.

⁷³ *Lewandowski*, Web Information Retrieval, 2005, S. 122.

⁷⁴ *Lewandowski*, Web Information Retrieval, 2005, S. 123.

⁷⁵ Als Ankertext bezeichnet man den anklickbaren Text (z. B. „Seite der Universität Passau“) einer Website, hinter dem sich ein weiterführender Link verbirgt, in HTML: ` Seite der Universität der Passau `.

verweist. Die inhärente Logik dabei ist, dass ein Ankertext in der Sache einer Fremdbeschreibung entspricht und mehr über den Inhalt einer Seite aussagt als diese oft selbst.⁷⁶

Als nutzerabhängige Variablen zur Gewichtung sind zum Beispiel die IP-Adresse, damit mittelbar der jeweilige näherungsweise Standort, oder das verwendete Betriebssystem zu nennen, schlussendlich auch das bisherige Suchverhalten (soweit bekannt) und sonstige Merkmale, die dem Feld der Personalisierung unterfallen.⁷⁷

Ergänzend können neben diesen gängigen Grundfaktoren auch beliebig weitere Faktoren aus der Indexdatenbank in die Berechnung eines Endwertes für die Relevanz einer Seite aufgenommen werden. Google zum Beispiel soll mehr als 200 verschiedene Kriterien innerhalb seines Suchalgorithmus verwenden.⁷⁸ Ob und in welchem Umfang der ursprüngliche Page Rank überhaupt noch Bedeutung für Google hat, ist unklar.⁷⁹ Es ist also in jedem Fall jenseits der gängigen Information-Retrieval-Modelle, wie eben der Textanalyse oder dem in den Grundzügen bekannten Page-Rank-Verfahren, von einer hohen individuellen Anpassung des Ranking-Verfahrens durch den jeweiligen Suchmaschinenanbieter auszugehen.⁸⁰

Zu den Gewichtungsprozessen der Suchmaschinen sei abschließend noch angemerkt, dass die dargestellten Modelle ihre Berechtigung vornehmlich aus ihrem Erfolg in der Praxis ziehen. Es handelt sich um eine dynamische, nie ganz abgeschlossene Entwicklung. Ein theoretisch identifizierbares Ideal existiert nicht.⁸¹ Die konkrete Gewichtung ist damit stets ein rein systemseitiger, anbieterabhängiger Versuch, Relevanz zu bewerten.⁸²

⁷⁶ Brin/Page, Computer Networks and ISDN Systems 1998, 107, 110.

⁷⁷ Siehe dazu unten Kapitel 4.4.

⁷⁸ Gillespie, in: Gillespie/Boczowski/Foot (Hrsg.), Media technologies, 2014, S. 175; die genaue Zahl und Aussagekraft der Zahl ist freilich unbekannt, vgl. ebda., dort Fn. 2, sie wird aber immer wieder genannt, vgl. exemplarisch Winston, Dalhousie Journal of Interdisciplinary Management 2011, 1, S. 4; Kühling/Gauß, ZUM 2007, 881, 883.

⁷⁹ Vgl. Lewandowski, Suchmaschinen verstehen, 2015, S. 101; als dokumentiertes Verfahren für Wertentscheidungen in der Gestaltung von Suchmaschinen bleibt er aber zumindest als Beispiel relevant; Google selbst führt ihn nach wie vor als wichtigen Bestandteil auf, vgl. <http://www.google.com/insidesearch/howsearchworks/algorithms.html> (Zugriff am 16.02.2016).

⁸⁰ Griesbaum/Bekavac/Rittberger, in: Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen, 2009, S. 36.

⁸¹ Lewandowski, Suchmaschinen verstehen, 2015, S. 89.

⁸² Lewandowski, Suchmaschinen verstehen, 2015, S. 30.

2.3 Algorithmen als Motor

Zentral für diese systemseitige Bewertung sind die verwendeten Algorithmen.⁸³ Ähnlich wie Suchmaschinen im Allgemeinen, ist auch der Begriff des Algorithmus mittlerweile in der gesellschaftspolitischen Diskussion um Suchmaschinen angekommen.⁸⁴ Das mathematische Begriffsverständnis von Algorithmen muss für die weiteren Ausführungen vorausgesetzt werden und wird im Folgenden dargestellt.

Die Arbeitsweise einer Suchmaschine – vom Crawling bis zur Abfrage – setzt grundsätzlich ihre automatische Ausführbarkeit durch Computersysteme voraus. Diese Ausführbarkeit wird durch Algorithmen gewährleistet.⁸⁵ Google beispielsweise gibt dazu an: „*Algorithmen sind Computerprogramme, die über Signale herausfinden, was die besten Ergebnisse für Ihre Anfrage sind*“.⁸⁶ Es wird damit eine eigene intelligente Entscheidungsfindung und echte Reaktion auf Fragestellungen suggeriert.

Nun mystifiziert diese Darstellung die praktische Realität. Mathematisch ist ein Algorithmus lediglich ein automatisierbares Verfahren, das in endlich vielen Schritten eine Eingabe zu einer Ausgabe verarbeitet. Die endlich vielen Schritte müssen zudem in endlicher Zeit abgeschlossen, eindeutig formuliert und ausführbar sein.⁸⁷ Hinzu kommen noch die Eindeutigkeit eines jeden Schrittes und das Vorhandensein von Variablen, die mit Elementen aus einer vorher zu bestimmenden Menge besetzt werden können.⁸⁸

Letztendlich liegt mit einem Algorithmus also nichts anderes als ein fester und endlicher Satz an Regeln zur Lösung eines bestimmten Problems vor. Er setzt möglichst umfangreiches Wissen über das Problem voraus und enthält inhärent Strategien zur Lösung desselbigen.⁸⁹ Grundsätzlich muss dabei zwischen dem Algorithmus und seiner Implementierung in durch Computersysteme ausführbaren Programmcode unter-

⁸³ Als Begriff vermutlich vom Namen des Mathematikers Al-Hwarizmi abgeleitet, vgl. *Barth*, Algorithmik für Einsteiger, 2003, S. 9.

⁸⁴ Beispielsweise *Lobe*, Brauchen wir noch Gesetze, wenn Rechner herrschen, FAZ Ausgabe Nr. 5 2015, S. 13.

⁸⁵ Algorithmen sind Voraussetzung wie Grenze der Fähigkeiten von Computersystemen, vgl. v. *Hellfeld*, GRUR 1989, 471, 477 f.

⁸⁶ Siehe <http://www.google.com/insidesearch/howsearchworks/algorithms.html> (Zugriff am 16.02.2016).

⁸⁷ *Barth*, Algorithmik für Einsteiger, 2003, S. 4.

⁸⁸ *Marly*, Urheberrechtsschutz für Computersoftware in der Europäischen Union, 1995, S. 113.

⁸⁹ *Introna*, Science, Technology & Human Values 2015, 1, 5.

schieden werden.⁹⁰ Weil eben Letzterer lediglich Regelsätze für Maschinen lesbar macht, können diese kaum als intelligent bezeichnet werden.⁹¹ Aber auch der Algorithmus selbst verfügt als präzise definierte Schrittfolge über keine Form von Eigenleben.⁹² Auch die Rechtsprechung⁹³ und Literatur⁹⁴ folgen diesem mathematischen Verständnis von Algorithmus.

2.4 Personalisierung

2.4.1 Personalisierung als Konzept

Suchmaschinen treten mit dem Ziel an, die Informationsfülle im Web zu organisieren und dem Nutzer möglichst relevante Ergebnisse für seine Suchanfrage zu liefern. Nun stehen sie dabei vor einem technologischen Grunddilemma. Relevanz ist ein auf dem Gebiet des Information Retrieval nicht allgemeingültig bestimmbares Merkmal. Letztendlich liefern Suchmaschinen damit zwangsnotwendig – in ihrer nicht personalisierten Form – homogene Ergebnisse für Nutzer mit ganz heterogenen Suchzielen.⁹⁵ Hinzu kommt auch, dass die Schlüsselwörter der Suchanfragen oft wenig geeignet sind, um eine echte Fragestellung des einzelnen Nutzers abzuleiten.⁹⁶

Als Lösungsweg erscheint nun die Berücksichtigung von als persönlich klassifizierten⁹⁷ Informationen im Rahmen des Ranking-Algorithmus. Man unternimmt den Versuch einer Personalisierung, verstanden als fortlaufende Anpassung von Inhalten und Struktur der Trefferliste an die spezifischen Bedürfnisse, Interessen und Vorlieben des einzelnen Nutzers.⁹⁸ Erkennt die Suchmaschine automatisiert den Nutzer hinter einer jeden Suchanfrage, ist präziser abzuschätzen, was dieser tatsächlich sucht und was für

⁹⁰ Vgl. *Introna*, Science, Technology & Human Values 2015, 1, 5.

⁹¹ *Barth*, Algorithmenik für Einsteiger, 2003, S. 11; v. *Hellfeld*, GRUR 1989, 471, 478 und insbesondere dort Fn. 46.

⁹² Vgl. v. *Hellfeld*, GRUR 1989, 471, 478, insb. dort Fn. 38.

⁹³ BGHZ 94, 276–292, Rn. 79 (zitiert nach *Juris*), dort allerdings missverständlich, weil auf einen „vorgegebenen Rechner bezogen“; grundsätzlich sind Algorithmen freilich rechnerunabhängig, vgl. v. *Hellfeld*, GRUR 1989, 471, 477, dort insb. Fn. 37.

⁹⁴ Vgl. *Hübner*, GRUR 1994, 883, 884; *Wiebe*, in: Spindler/Schuster (Hrsg.), Recht der elektronischen Medien, 32015, § 69a UrhG, Rn. 22; *Golla*, PinG 2014, 61; grundlegend *Marly*, Urheberrechtsschutz für Computersoftware in der Europäischen Union, 1995, S. 112 ff.

⁹⁵ *Goldmann*, 8 Yale Journal of Law and Technology 188 (2006), 198.

⁹⁶ Zum grundlegenden Problem des Query Understanding weiterführend *Lewandowski*, in: ders. (Hrsg.), Handbuch Internet-Suchmaschinen – 2, 2011, S. 55 ff.

⁹⁷ Dazu im Folgenden, vgl. auch Fn. 100.

⁹⁸ Vgl. *Elixmann*, Datenschutz und Suchmaschinen, 2012, S. 44 m. w. N.; mit einer grafischen Darstellung *Bozdag*, Ethics and Information Technology 2013, 209, 215.

ihn interessant sein wird; so jedenfalls das ideale Szenario.⁹⁹ In diesem Sinne soll Personalisierung hier begrifflich weit verstanden werden. Auch die Berücksichtigung der Sprache oder des näherungsweise bestimmten Aufenthaltsortes fallen darunter. Ohne Berücksichtigung bleibt hier die Frage nach dem Personenbezug der möglicherweise erfassten Informationen im Sinne des Datenschutzrechts.¹⁰⁰

Neben technologischen Vorteilen bietet dieser Ansatz auch wirtschaftliche Anreize.¹⁰¹ Das Werbegeschäft ist noch immer absolut primäre Einnahmenquelle der Unternehmen hinter den Suchmaschinen.¹⁰² Personalisierung bietet die Möglichkeit, passgenauere Werbeanzeigen zu schalten, und damit entsprechende Vorteile bei der Werbevermarktung.¹⁰³

Der Umfang jener Daten, die als persönlich klassifiziert und vom Ranking-Algorithmus berücksichtigt werden, ist äußerst variabel und mit Blick auf die Praxis konkreter Angebote unbekannt; auch hier greift das Phänomen der Suchmaschine als Black-box.¹⁰⁴

Die bekannteste Möglichkeit zur Datensammlung ist der Web-Cookie, eine Text-Datei, die auf dem PC des Nutzers abgelegt wird und von der Suchmaschine ausgelesen werden kann.¹⁰⁵ In ihr lassen sich zahlreiche Informationen speichern, zum Beispiel der bisherige Suchverlauf oder die Frage, welche Suchtreffer angeklickt wurden. Daneben lassen sich Geräteinformationen wie das Betriebssystem oder der User-Agent des Browsers, die IP-Adresse – vor allem auch als Merkmal zur Geolokalisation – sowie der Browserverlauf heranziehen. Auch Informationen, welche aus mit Google assoziierten Konten gewonnen werden, können nützlich sein. Bei Letzteren ist eine

⁹⁹ Idealisierend und stark für Personalisierung *Goldmann*, 8 *Yale Journal of Law and Technology* 188 (2006), 198 f.

¹⁰⁰ Zum datenschutzrechtlichen Personenbezug siehe unten Kapitel 8.3.3.2.2.

¹⁰¹ *Rotenberg*, in: Machill (Hrsg.), *Die Macht der Suchmaschinen*, 2007, S. 89.

¹⁰² Vgl. <http://investor.google.com/financial/tables.html> (Zugriff am 16.02.2016); *Winston*, *Dalhousie Journal of Interdisciplinary Management* 2011, 1, S. 6; eine Ausnahme bildet insofern Microsoft mit seiner Suchmaschine Bing, aber auch dort dürfte eine langfristige Querfinanzierung aus dem Microsoft-Kerngeschäft nicht gewünscht sein.

¹⁰³ *Feuz/Fuller/Stalder*, 16 *First Monday* (2011), abrufbar unter <http://journals.uic.edu/ojs/index.php/fm/article/view/3344> (Zugriff am 16.02.2016); *Rotenberg*, in: Machill (Hrsg.), *Die Macht der Suchmaschinen*, 2007, S. 89.

¹⁰⁴ Vgl. *Hannak/Sapiezynski/Kakhki u. a.*, in: Schwabe (Hrsg.), *Proceedings of the 22nd International Conference on the World Wide Web*, 2013, S. 528; wie erwähnt soll Google auf über 200 Signale zurückgreifen.

¹⁰⁵ Vgl. *Zeidler/Brüggemann*, CR 2014, 248, 250 ff.; weiterführend zum Web-Tracking *Roesner/Kohn/Wetherall*, *Detecting and Defending against Third-Party Tracking on the Web*, 2012.

Auswertung statischer Informationen (Alter, Geschlecht), aber auch eine Profilbildung anhand (vermuteter) sozialer Interessen und Vorlieben denkbar.¹⁰⁶ Ein Fall allgemein bekannter Personalisierung im weitesten Sinne wäre die Berücksichtigung der Sprache des Nutzers. Ein relativ neuer Ansatz macht sich gerade Mechanismen des Selbstschutzes der Nutzer zu eigen. Je stärker Nutzer ihr technisches System individualisieren, so auch durch jene Browser-Erweiterungen, welche gerade eine Personalisierung verhindern sollen, desto höher wird der Wiedererkennungswert. Ein „Fingerabdruck“ ihres Browsers entsteht, der sich speichern und verwenden lässt.¹⁰⁷

2.4.2 Personalisierung in der gegenwärtigen Praxis

Nach Darstellung der Personalisierung aus theoretischer Sicht soll kurz das Realbeispiel Google in den Blick genommen werden.

Google stellte erstmals 2004 eine personalisierte Suchfunktion zur Verfügung, damals noch auf experimenteller Basis und neben der normalen Suchfunktion stehend.¹⁰⁸ Es folgte eine schrittweise Ausweitung auf die Suche von Nutzern, die über Nutzerkonten bei Google-Diensten verfügen. Seit 2009 ist die personalisierte Suche Teil der normalen Suchfunktion. Es gibt also – streng genommen – in den Standardeinstellungen keine nicht-personalisierte Suche über Google mehr.¹⁰⁹

Mit seiner neuen Datenschutzerklärung ging das Unternehmen schließlich 2012 noch einen Schritt weiter: Im Rahmen seiner internen Datenverarbeitung behandelt Google Inc. seitdem alle seine Dienste als einheitliches Netzwerk. Meldet man sich zum Beispiel bei einem Dienst an, ist man effektiv mit allen verbunden.¹¹⁰ Es kann somit der Weg des Nutzers durch alle Dienste nachvollzogen werden. Über sein weites Werbe-

¹⁰⁶ Vgl. *Lewandowski*, Suchmaschinen verstehen, 2015, S. 119, der ein frühes Personalisierungspatent aus dem Jahr 2003 zitiert; vgl. auch *Hannak/Sapiezynski/Kakhki u. a.*, in: Schwabe (Hrsg.), Proceedings of the 22nd International Conference on the World Wide Web, 2013, S. 532 f. Eine abschließende Darstellung theoretischer Personalisierungsmöglichkeiten kann im Rahmen dieser Arbeit nicht geleistet werden.

¹⁰⁷ *Zeidler/Brüggemann*, CR 2014, 248, 252.

¹⁰⁸ Zu den Anfängen: *Türker*, The optimal design of a search engine from an agency theory perspective, 2004, S. 24 f.

¹⁰⁹ *Hannak/Sapiezynski/Kakhki u. a.*, in: Schwabe (Hrsg.), Proceedings of the 22nd International Conference on the World Wide Web, 2013, S. 528.

¹¹⁰ *Funke*, CR 2012, R26-R27.

netzwerk ist theoretisch eine umfangreiche Rekonstruktion des Surfverhaltens eines Nutzers über den Kreis der eigenen Dienste hinaus möglich.¹¹¹

Welchen Umfang nun die selbsterklärte Personalisierung auf die gegenwärtigen Ergebnislisten hat, ist derzeit noch nicht abschließend geklärt.¹¹² Eine Studie beispielsweise, welche Reverse-Engineering-Ansätze bemüht, zeigt, dass Personalisierung stattfindet. Im Falle Googles ist diese aber nur für Suchanfragen messbar, welche bei gleichzeitiger Verwendung eines Google-Kontos erfolgen. Einen gravierenden Einfluss hat weiterhin die IP-Adresse, welche der Geolokalisation dient.¹¹³

Der gravierende Einfluss der IP-Adresse wird durch eine weitere Studie gestützt. Diese bemüht sich allerdings um einen Versuchsaufbau, der tendenziell eher dem Suchverhalten realer Nutzer entsprechen soll, und kommt zu einem weit stärkeren Einfluss der Suchhistorie auf die angezeigten Trefferlisten.¹¹⁴ Im Gegensatz hierzu kam eine andere Studie, die speziell eine potentielle Reduktion der Ergebnisvielfalt und einen Perspektivenwechsel bei den dargestellten Treffern erforschen wollte, zum Ergebnis, dass – jedenfalls im Testzeitraum – keine signifikante Personalisierung nachweisbar war. Gleichzeitig verweist diese Studie allerdings auch auf methodische Schwierigkeiten und Problematiken der Blackbox.¹¹⁵

Zur Praxis der Personalisierung ergibt sich damit ein höchst offenes Bild. Es lässt sich einerseits bereits jetzt eine Abkehr von universellen Trefferlisten für jedermann, nur abhängig von den Suchbegriffen, feststellen. Gleichzeitig bleibt der empirisch fassbare und mehrfach reproduzierbare Beleg für umfangreiche Personalisierung zum jetzigen Zeitpunkt aus. Hinzu kommt allerdings das Paradoxon, dass eine empirisch angelegte Forschung zwar wünschenswert wäre, allerdings stets unter dem methodischen Vorbe-

¹¹¹ Hannak/Sapiezynski/Kakhki u. a., in: Schwabe (Hrsg.), Proceedings of the 22nd International Conference on the World Wide Web, 2013, S. 528.

¹¹² Siehe zur Problematik Kapitel 2.1.2. Die Annahme von Bloch, Meinungsvielfalt contra Medienmacht, 2013, S. 258, eine umfassende Personalisierung sei gegeben, ist dementsprechend wohl noch verfrüht.

¹¹³ Hannak/Sapiezynski/Kakhki u. a., in: Schwabe (Hrsg.), Proceedings of the 22nd International Conference on the World Wide Web, 2013, S. 536.

¹¹⁴ Xing/Meng/Doozan u. a., in: Faloutsos/Kuzmanovic (Hrsg.), Passive and active measurement, 2014, S. 138 f.

¹¹⁵ Jürgens/Stark/Magin, in: Stark/Dörr/Aufenanger (Hrsg.), Die Googleisierung der Informationssuche, 2014, S. 128 f.

halt der Blackbox steht.¹¹⁶ Andererseits finden sich Trendprognosen – auch und gerade von Suchanbietern selbst –, die in der Personalisierung sowohl finanziell als auch qualitativ die Zukunft sehen.¹¹⁷ Auch ist das gesamte Themenfeld Big Data, das gemeinhin als zentraler Gegenstand zukünftiger Informationstechnologie gesehen wird, von der Erfassung von Daten zur Personalisierung geprägt.¹¹⁸

2.5 Zusammenfassung zum Untersuchungsgegenstand

Zusammengefasst lässt sich – um die eingangs angeführte Definition aufzugreifen und zu ergänzen – festhalten:

Suchmaschinen sind Computersysteme, die verteilte Inhalte (Dokumente) aus dem World Wide Web mittels Crawling erfassen und nach internen Kriterien indexieren, um sie anschließend über eine Benutzerschnittstelle – die Website des Suchanbieters – für die Nutzer durchschaubar zu machen. Die Trefferliste stellt dabei eine Reihung dar, deren Gewichtung nach anbieter eigenen Kriterien und mittels vorgegebener Regelsätze (Algorithmen) erfolgt. Die Algorithmen bilden dabei die internen Gewichtungen ab und machen sie automatisiert ausführbar. Um systemseitige Relevanzvorstellungen stärker um (vermutete) Nutzervorstellungen zu ergänzen, erscheint eine weitergehende Personalisierung jedenfalls als Zukunftstrend.

¹¹⁶ So weisen *Jürgens/Stark/Magin* zutreffend darauf hin, dass anbieterseitig Forschungsvorhaben durchaus erkannt und entsprechend beeinflusst werden könnten, *Jürgens/Stark/Magin*, in: Stark/Dörr/Aufenanger (Hrsg.), *Die Googleisierung der Informationssuche*, 2014, S. 130; zum Thema auch *Lazer*, *Science* 2015, 1090.

¹¹⁷ Vgl. <http://www.heise.de/newsticker/meldung/Suchmaschinen-Google-und-Microsoft-bauen-am-ultimativen-Assistenten-2579572.html> (Zugriff am 16.02.2016).

¹¹⁸ Aus (rechts)politischer Perspektive *Mayer-Schönberger/Cukier*, *Big data*, 2013.

Suchmaschinen, Algorithmen und Meinungsmacht
Eine verfassungs- und einfachrechtliche Betrachtung

Hartl, K.

2017, XII, 276 S., Softcover

ISBN: 978-3-658-16826-1