

Dirk Lewandowski

## D 1 Suchmaschinen

### D 1.1 Einleitung

Eine Suchmaschine (auch: Web-Suchmaschine; Universalsuchmaschine) ist ein Computersystem, das Inhalte aus dem *World Wide Web* mittels *Crawling* erfasst und über eine Benutzerschnittstelle durchsuchbar macht, wobei die Ergebnisse in einer nach systemseitig angenommener Relevanz geordneten Darstellung aufgeführt werden.

Dies bedeutet, dass Suchmaschinen im Gegensatz zu anderen Informationssystemen nicht auf einem klar abgegrenzten Datenbestand aufbauen, sondern sich ihre Inhalte durch Traversieren des Web zusammensuchen. Man kann Suchmaschinen daher auch als Metainformationsdienste bezeichnen, welche nicht auf einen selbst erstellten Datenbestand zurückgreifen, sondern durch Erfassung der Inhalte unterschiedlicher Datenangebote aus dem Web den eigenen Datenbestand (*Index*) aufbauen. Zum anderen grenzt diese Definition Suchmaschinen von anderen Suchsystemen, die nicht auf Inhalte des *World Wide Web* zurückgreifen, ab.

Suchmaschinen sind zuerst einmal Information-Retrieval-Systeme, deren Besonderheit vor allem darin liegt, dass sie eben ihre Indices aus den verstreut vorliegenden Dokumenten des *World Wide Web* aufbauen und sich außerdem hinsichtlich der Benutzerführung und Trefferdarstellung dem Zielpublikum der Laiennutzer anpassen. Mittlerweile hat sich das Verhältnis zwischen Suchmaschinen und konventionellem Information Retrieval aber zunehmend umgekehrt: Während frühe Suchmaschinen sich noch sehr stark an konventionellen Information-Retrieval-Systemen orientierten, kommen viele der wesentlichen Entwicklungen des Information Retrieval in den letzten Jahren entweder direkt von den bekannten Suchmaschinenanbietern oder entstanden in Kooperation mit ihnen. So sind die großen Suchmaschinenanbieter auch Vorreiter für Entwicklungen im Information Retrieval geworden. Andere Informationssysteme orientieren sich vor allem in Hinblick auf die Benutzerführung und die Trefferanordnung und -darstellung an ihnen. Beispielhaft kann auch angeführt werden, dass Information-Retrieval-Lehrbücher sich zunehmend auf Suchmaschinen beziehen bzw. den Anwendungsfall Suchmaschinen bereits in ihrem Titel tragen (Lit. 01, Lit. 02, Lit. 03).

#### D 1.1.1 Zugänge zu den Inhalten im World Wide Web

Suchmaschinen sind bei weitem nicht der einzige Zugang zu Daten im Web. Wenn man jedoch die Nutzungszahlen betrachtet, kann man feststellen, dass sie sich weitgehend als *der* Zugang zu den Angeboten im Web etabliert haben. Die Popularität der Suchmaschinen geht sogar so weit, dass sie neben E-Mail der beliebteste Dienst des Internet sind (Lit. 04). Die wichtigsten alternativen Zugänge zu Web-Inhalten (vgl. auch Lit. 05) sind:

- Web-Verzeichnisse (auch: Web-Kataloge; Lit. 06), welche *Websites* klassifizieren, was ein hierarchisches Browsing ermöglicht und vor allem für die Suche nach einigen guten Quellen zu einem Thema mit hoher *precision* geeignet ist. Während Web-Verzeichnisse vor allem in den frühen Jahren des *World Wide Web* beliebt waren, haben sie heute nur noch geringe Bedeutung.
- *Social-Bookmarking*-Dienste (vgl. Lit. 07), in denen Nutzer Webseiten bzw. *Websites* speichern, annotieren und mit anderen Nutzern teilen können.
- Frage-Antwort-Dienste (vgl. Lit. 08), in denen Nutzer der Nutzer-Gemeinschaft Fragen stellen können, die asynchron beantwortet werden können.

Einen Sonderfall stellen Metasuchmaschinen dar, die auf Nutzerseite den allgemeinen Suchmaschinen vergleichbar sind, jedoch selbst keinen Index aufbauen, sondern auf die Indices mehrerer anderer, *echter* Suchmaschinen zurückgreifen. Dabei wird eine Suchanfrage in dem Moment, in dem ein Nutzer sie an die Meta-Suchmaschine stellt, an die zugrunde liegenden Suchmaschinen weitergeleitet. Die zurückgegebenen Ergebnisse werden zusammengefasst und in einer einzigen Trefferliste zusammengefasst. Die Vorteile von Metasuchmaschinen werden in einer größeren Abdeckung des Webs und in einer besseren Relevanzbewertung durch die Zusammenfassung der *Urteile* der zugrundeliegenden Suchmaschinen gesehen. Allerdings sind diese Annahmen nicht empirisch bestätigt. In der Praxis spielen Metasuchmaschinen keine Rolle mehr, und auch die theoretische Fundierung des Ansatzes wird stark angezweifelt (Lit. 09).

Von den allgemeinen Suchmaschinen sind die Spezialsuchmaschinen zu unterscheiden (vgl. Lit. 10). Diese sind thematisch beschränkt und machen damit eine zielgenaue Recherche möglich, wobei das Ranking speziell auf die von ihnen erschlossenen Dokumente angepasst werden kann, ebenso wie die sachliche Erschließung der Dokumente. Schließlich ergeben sich auch Vorteile bei der Darstellung der Ergebnisse, welche auf den individuellen Zweck der Spezialsuchmaschinen sowie auf das Niveau der Zielgruppe angepasst werden kann (Lit. 10, S. 56). Dass Spezialsuchmaschinen nicht durch Universalsuchmaschinen ersetzt werden können, ergibt sich aus den Problemen der letzteren:

- Sie haben technische Beschränkungen und können (trotz des Etiketts *universal*) nicht das gesamte Web abdecken,
- es bestehen finanzielle Hürden, die die Erfassung der Inhalte und ihre Erschließung beschränken,
- Universalsuchmaschinen richten sich an den *Durchschnittsnutzern* aus,
- sie müssen eine einheitliche Erschließung aller indexierten Inhalte betreiben.

### D 1.1.2 Anfragetypen

Um die große Zufriedenheit der Nutzer mit Suchmaschinen zu verstehen, aber auch, um die Problemfelder bei der Websuche aufzeigen zu können, ist vor allem ein Verständnis der an die Suchmaschinen gerichteten Anfragen wichtig, und hier vor allem die Unterscheidung nach Anfragetypen. Von Andrei Broder (Lit. 11) stammt die maßgebliche Unterscheidung nach informationsorientierten (*informational*), navigationsorientierten (*navigational*) und transaktionsorientierten (*transactional*) Suchanfragen.

- Mit navigationsorientierten Anfragen soll eine Seite (wieder)gefunden werden, die dem Benutzer bereits bekannt ist oder von der er annimmt, dass sie existiert. Beispiele sind die Suche nach Homepages von Unternehmen (*Microsoft*) oder nach Personen (*Heidi Klum*). Solche Anfragen haben in der Regel ein richtiges Ergebnis. Das Informationsbedürfnis ist befriedigt, sobald die gewünschte Seite gefunden wird.
- Bei informationsorientierten Anfragen ist das Informationsbedürfnis meist nicht durch ein einziges Dokument zu befriedigen. Der Nutzer möchte sich stattdessen über ein Thema informieren und liest deshalb mehrere Dokumente. Informationsorientierte Anfragen zielen auf jeden Fall auf statische Dokumente, nach dem Aufruf des Dokuments ist also keine weitere Interaktion auf der Website nötig, um an die gewünschten Informationen zu gelangen.
- Mit transaktionsorientierten Anfragen wird eine Website gesucht, auf der anschließend eine Transaktion stattfindet, etwa der Kauf eines Produkts, der Download einer Datei oder die Recherche in einer Datenbank.

Die Angaben über die Anteil der Anfragetypen am Gesamt der Suchanfragen schwankt, da Studien, die sich mit diesem Thema beschäftigen, nicht auf einer gemeinsamen Datenbasis beruhen und sich hinsichtlich verwendeter Suchmaschine, Zeitraum und Klassifikationsmethode unterscheiden

(vgl. Lit. 12). Allerdings sind sich die Studien einig, dass alle drei Anfragetypen einen nennenswerten Anteil der Anfragen ausmachen. Insbesondere der hohe Anteil der navigationsorientierten und der einfachen informationsorientierten Anfragen legt die Vermutung nahe, dass sich die Zufriedenheit der Nutzer mit den Suchmaschinen zu einem großen Teil durch die erfolgreiche Beantwortung dieser Suchanfragen erklären lässt.

### D 1.1.3 Suchmaschinenmarkt

Der weltweite Markt für Web-Suchmaschinen wird momentan von nur wenigen Anbietern bestimmt. Allen voran ist hier die Suchmaschine Google zu nennen, welche weltweit einen Marktanteil (gemessen in Suchanfragen) von mehr als zwei Dritteln hat (Lit. 13). Stärkster Konkurrent zu Google ist Microsoft mit seiner Suchmaschine Bing. Auch wenn diese keine vergleichbaren Werte erreichen kann, so erreicht sie doch zumindest in den USA einen recht hohen Marktanteil von etwa 16 Prozent (zusammen mit seinem Partner Yahoo, der Bing-Ergebnisse anzeigt, knapp 29 Prozent; Lit. 14), während sie international eine weit geringere Rolle spielt. In Europa sind die nationalen Suchmaschinenmärkte durch ein Quasi-Monopol von Google gekennzeichnet (Lit. 15). Diese Suchmaschine erreicht in den meisten Ländern einen Anteil von mehr als 90 Prozent der Suchanfragen. Außerhalb Europas haben sich teilweise gewichtige nationale Suchmaschinen etablieren können; etwa Baidu in China oder Yandex in Russland. Inwieweit es diesen Anbietern gelingen wird, sich auch international zu platzieren, bleibt abzuwarten.

Zu unterscheiden ist zwischen Suchmaschinen, die einen eigenen Datenbestand nutzen (die bisher genannten) und solchen, die auf den Index eines Partners zurückgreifen und dadurch eigentlich nicht als eigenständige Suchmaschinen bezeichnet werden können. Prominentestes Beispiel hierfür ist Yahoo, welches seine eigene Suchmaschine im Jahr 2009 aufgegeben hat zugunsten einer Kooperation mit Bing, dessen Ergebnisse seitdem unter dem Yahoo-Label angezeigt werden. In Deutschland greifen beispielsweise Portale wie T-Online, Web.de, GMX und AOL auf Google-Ergebnisse zurück. Rechnet man die Marktanteile dieser Suchportale, welche sämtlich auf Google zurückgreifen, zu dem Marktanteil von Google hinzu, so ergibt sich für diese Suchmaschine ein Marktanteil von mehr als 90 Prozent (Lit. 16). Auf die daraus entstehenden Probleme wird weiter unten eingegangen.

Zu unterscheiden ist damit zwischen einem Markt für Suchmaschinen als Produkte, die sich an den Endkunden wenden (und oft nicht viel mehr sind als ein Suchinterface, das auf einen der bekannten Indices zurückgreift) und einem Markt für Web-Indices, die primär durchaus auch für andere Zwecke als für die Suche erstellt sein können (hier sind vor allem Analysezwecke im Bereich Suchmaschinenoptimierung sowie die Reputationsanalyse für Unternehmen zu nennen).

## D 1.2 Aufbau algorithmischer Suchmaschinen

In diesem Abschnitt wird beschrieben, wie eine Suchmaschine grundlegend aufgebaut ist. Empfehlenswerte und umfangreiche Darstellungen finden sich in Lit. 02 und Lit. 03. Neben der Datenbasis (dem Index) besteht eine Suchmaschine aus drei Hauptkomponenten: Dem *Crawler*, dem *Indexer* und dem *Searcher*. Die Unterteilung der Bereiche folgt Risvik & Michaelsen (Lit. 17).

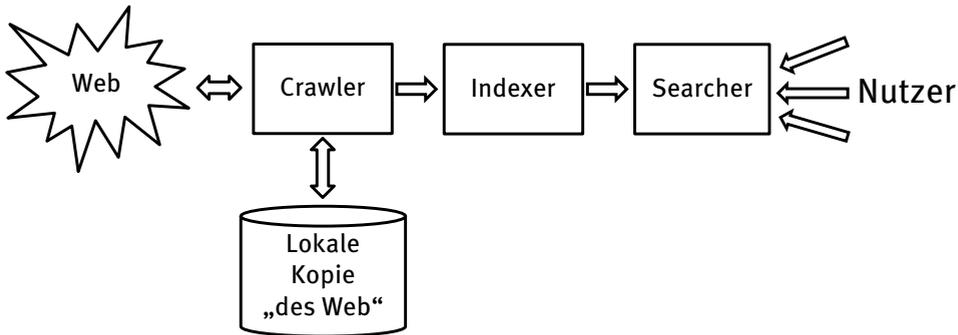


Abb. 1: Aufbau einer Web-Suchmaschine (nach Lit.17)

### D 1.2.1 Datenbasis

Als Datenbasis dienen den Suchmaschinen die Inhalte des Webs, die eigenständig aufgefunden werden müssen. Da die Inhalte im Web verteilt vorliegen, kann eine Vollständigkeit des Datenbestands, der im Idealfall eine aktuelle und vollständige *lokale Kopie des Web* ist, nicht gewährleistet werden.

Web-Suchmaschinen haben den Anspruch, einen möglichst hohen Anteil des gesamten Web zu erfassen und für die Suche verfügbar zu machen. Die damit verbundenen Schwierigkeiten werden deutlich, wenn man die Menge der im Web vorhandenen Dokumente betrachtet. Diese kann zwar nicht exakt ermittelt werden, und eine solche Zahl wäre aufgrund der nahezu beliebigen automatischen Kombinierbarkeit von Inhalten zu immer neuen Dokumenten auch nicht aussagekräftig, aber Suchmaschinen sehen sich doch vor der Herausforderung, einen Index aus vielen Milliarden Dokumenten zusammensetzen zu müssen. Deren Indexierung ist nicht nur ressourcenintensiv und kostspielig, sondern wird durch zwei Faktoren erschwert: Zum einen muss der Index kontinuierlich aktuell gehalten werden, zum anderen müssen sogenannte Spam-Dokumente aus dem Index ausgeschlossen werden. Dabei handelt es sich um von der Suchmaschine nicht erwünschte Dokumente, vor allem um solche, die einzig und allein zu dem Zweck erstellt wurden, Suchmaschinen und Nutzer über ihre tatsächliche Intention zu täuschen.

Während Suchmaschinen früher einen einzigen Index (den Web-Index) aufbauten, aus dem alle Suchanfragen bedient wurden, haben sich inzwischen verschiedene Standard-Indices etabliert, aus denen eine Web-Suchmaschine besteht. Dazu gehören neben dem Web-Index ein Nachrichten-, ein Bilder- und ein Video-Index.

Während es der Anspruch des Web-Index ist, das Web möglichst vollständig abzudecken, fokussieren die Spezialsuchen (vertikale Indices) auf bestimmte Themen und werden aus begrenzten Quellenmengen zusammengestellt, welche vorab ausgesucht wurden. So basieren Nachrichtensuchmaschinen auf den Dokumenten aus zuvor intellektuell ausgewählten Quellen. Durch die Beschränkung der vertikalen Suchen auf klar umgrenzte Kollektionen lassen sich Indexierungsintervalle und -tiefe den entsprechenden Bedürfnissen anpassen. So ist zum Beispiel bei einer Nachrichtensuchmaschine aufgrund des hohen Aktualitätsanspruchs eine häufigere Indexierung nötig (und aufgrund der weit geringeren Dokumentenmenge auch möglich) als bei der Erstellung des allgemeinen Web-Index.

In Tab. 1 sind beispielhaft einige Kollektionen der Suchmaschine Google aufgezählt und kurz erläutert. Selbstverständlich können alle diese Kollektionen direkt angesteuert werden (damit handelt es sich um Spezialsuchmaschinen, s.o.), allerdings werden auch Treffer aus allen diesen Kollektionen in die Trefferlisten der Websuche eingebunden (*Universal Search*). Dadurch ergibt sich eine besondere Vielfalt, die durch die Verwendung eines einzigen Index nicht zu erreichen wäre.

Bei den vertikalen Kollektionen zeigt sich, dass nicht (nur) wie in der Websuche auf die allgemein im Web zugänglichen Inhalte zurückgegriffen wird (oder eine Auswahl daraus; siehe Lit. 10, S. 57f), sondern auch gesonderte Kollektionen aufgebaut werden (bspw. durch das Einscannen von gedruckten Büchern für die Google Buchsuche). Vaidhyanathan (Lit. 18) teilt die Dienste von Google in drei Bereiche ein, die auch auf andere Suchmaschinen übertragbar sind:

1. Scan and link: Fremde Inhalte werden erfasst, aggregiert und zur Suche zur Verfügung gestellt (Beispiel: Websuche).
2. Host and serve: Von den Nutzern selbst erstellte Inhalte werden auf der eigenen Plattform gesammelt und gehostet (Beispiele: Blogger, Youtube).
3. Scan and serve: Google bringt Dinge aus der echten Welt in die digitale (Beispiele: Google Books, Google Street View)

Von den allgemeinen Suchmaschinen, die vordergründig nur den ersten Punkt bedienen, wird erwartet, dass sie in allen drei Bereichen aktiv sind. Aus dem eigenen Hosten von Inhalten (Punkt 2) ergibt sich der Vorteil, dass umfangreiche Metadaten (von der Autorenverifikation bis zu Nutzungsdaten) anfallen und die den Dienst betreibende Suchmaschine exklusiv über diese zusätzlichen Daten verfügt.

Bereich	URL	Erläuterung
Nachrichten	<a href="http://news.google.de">http://news.google.de</a>	Durchsucht werden die Inhalte von manuell ausgewählten Nachrichtenquellen. Bei der Indexerstellung besonderer Schwerpunkt auf der Aktualisierung des Index.
Bilder	<a href="http://www.google.de/imghp">http://www.google.de/imghp</a>	Durchsucht Bilder, die im freien Web gefunden wurden.
Maps	<a href="http://maps.google.de">http://maps.google.de</a>	Verbindet proprietäres Kartenmaterial mit Inhalten aus dem freien Web und Bewertungen aus dem eigenen Dienst Google Places.
Bücher	<a href="http://books.google.de">http://books.google.de</a>	Durchsucht eingescannte Bücher; Unterscheidung zwischen gemeinfreien Büchern, die im Volltext angepasst werden, und geschützten Werken, von denen ausgewählte Seiten angezeigt werden.
Shopping	<a href="http://www.google.de/shopping">http://www.google.de/shopping</a>	Produktsuche, die vor allem auf von den Händlern angelieferten strukturierten Daten beruht.
Videos	<a href="http://www.google.de/videohp">http://www.google.de/videohp</a>	Durchsucht Videos, die im Web auf verschiedenen Plattformen frei verfügbar sind.
Wissenschaft	<a href="http://scholar.google.de">http://scholar.google.de</a>	Durchsucht wissenschaftliche Artikel und Bücher, die entweder in Repositories, auf den Websites von Wissenschaftlern und Hochschulen oder durch kooperierende Verlage zur Verfügung gestellt werden. Verbindung von kostenlosen Inhalten/Versionen und kostenpflichtigen Werken.

**Tab. 1:** Kollektionen einer Suchmaschine (Beispiel Google; Auswahl)

### D 1.2.2 Crawling

Die Aufgabe des *Crawlers* (auch: *Spider*) ist es, neue Dokumente aufzufinden, indem Hyperlinks innerhalb bereits bekannter Dokumente verfolgt werden. Der *Crawling*-Vorgang findet kontinuierlich statt.

Da das Web keine zentrale Struktur hat und damit auch nicht auf eine bestimmbare absolute Menge von Dokumenten zugegriffen werden kann, müssen Suchmaschinen neue und aktualisierte Dokumente auf anderem Wege finden. Dies geschieht vor allem durch das Verfolgen von Hyperlinks. Als Basis dafür dient eine Ausgangsmenge (*seed set*) bekannter Webseiten. Die in diesen enthaltenen Links werden verfolgt, die so gefundenen Dokumente indexiert und wiederum die darin enthaltenen Links verfolgt. Auf diese Weise sollen möglichst alle im Web vorhandenen Dokumente gefunden werden. Dies ist jedoch in der Praxis nicht der Fall: Zum einen hängt die so erreichbare Abdeckung stark von dem *seed set* ab, zum anderen haben Untersuchungen gezeigt, dass die Verbindungen der Dokumente untereinander nicht so stark sind, dass durch einfache Linkverfolgung alle Dokumente zu finden wären (Lit. 19).

Neben dem Auffinden neuer Dokumente ist es die Aufgabe des *Crawlers*, bereits bekannte Dokumente auf Aktualisierungen zu prüfen und sicherzustellen, dass die Dokumente zwischenzeitlich nicht gelöscht wurden. *Crawling* ist also auch in Bezug auf die bereits bekannten Dokumente kein abgeschlossener Vorgang, sondern muss kontinuierlich stattfinden.

Suchmaschinen priorisieren die zu besuchenden Dokumente (deren URLs bereits bekannt sind und im *Crawler Queue* zwischengespeichert werden), beispielweise nach Popularität (gemessen in der Regel anhand der Verlinkungsstruktur, aber auch anhand von Zugriffen) oder nach Aktualisierungsintervall (häufig aktualisierte Website wie beispielweise Nachrichten-Websites werden häufiger auf Aktualisierungen und neue Dokumente geprüft).

*Crawler* können nur die Inhalte des Webs erreichen, die über Links zugänglich sind. Dabei handelt es sich um das *surface web* (Oberflächenweb), im Gegensatz dazu bezeichnet das *invisible web* (unsichtbares Web; auch: *deep web*) diejenigen Inhalte, auf die die Suchmaschinencrawler nicht zugreifen können. Gründe hierfür sind vor allem die fehlende Verlinkung, die Abschottung der Inhalte hinter Passwort-Abfragen und dynamische Inhalte, die erst im Moment einer Abfrage aus Datenbanken generiert werden (Lit. 20, Lit. 21, Lit. 22).

Inzwischen haben sich die Suchmaschinen auch darauf eingestellt, Inhalte aus dem *invisible web* zu indexieren. Spezielle *Crawler* stellen Suchanfragen in vorher gefundenen Suchformularen (Lit. 23). Dabei versuchen sie, passende Suchanfragen aufgrund des Kontexts des Suchformulars zu erschließen. Es geht dabei allerdings vor allem um das Auffinden von umfangreicheren Dokumenten (zum Beispiel im PDF-Format), und es ist bislang unklar, zu welchem Grad die Suchmaschinen mit dieser Methode das *invisible web* sichtbar machen.

Für den Aufbau der oben beschriebenen Kollektionen haben die Suchmaschinen jeweils eigene *Crawler* entwickelt, die bestimmte Datenbestände abfragen. So ist beispielweise ein *Crawler*, der Nachrichteninhalte für die News-Suche abfragen soll, anders eingestellt als der allgemeine *Web-Crawler*. Da die Nachrichtensuche nur eine recht beschränkte Menge ausgesuchter Quellen umfasst, ist es hier möglich, diese in sehr kurzen Abständen auf neue und veränderte Inhalte zu prüfen. *Crawler* für Spezialsuchmaschinen oder spezielle Kollektionen können auch so eingestellt werden, dass sie bestimmte Inhalte, die sonst zum *invisible web* gehören, mit abfragen können. So ist es beispielsweise dem *Crawler* von *Google Scholar* (durch Vereinbarungen mit den entsprechenden Verlagen) möglich, auch Inhalte abzufragen, die hinter der *Paywall* der Anbieter stehen.

### D 1.2.3 Indexer

Die Aufgabe des *Indexers* ist es, die vom *Crawler* gelieferten Dokumente so zu zerlegen und aufzubereiten, dass sie in der Suche effizient verarbeitet werden können. Das System zur Syntaxanalyse (*parsing module*) zerlegt die gefundenen Dokumente in indexierbare Einheiten (einzelne Wörter, Wortstämme oder N-Gramme) und verzeichnet deren Vorkommen innerhalb des Dokuments. Dadurch wird ein invertierter Index erstellt, der für jede indexierte Einheit diejenigen Dokumente verzeichnet, in denen diese vorkommt. So wird in der Suche ein schneller Zugriff ermöglicht, da nicht alle Dokumente durchsucht werden müssen, sondern nur anhand der Suchbegriffe abgeglichen

werden muss, in welchen Dokumenten diese vorkommen. Ein Nachteil des Verfahrens liegt darin, dass Informationen, die in diesem Prozess nicht berücksichtigt werden (beispielsweise vom Betreiber einer Website erstellte, umfangreiche Metadaten), später auch nicht recherchiert werden können.

Diese Form der Indexierung ist nicht suchmaschinenspezifisch (vgl. B 11 Automatische Indexierung und B 15 Modelle im Information Retrieval sowie aus KSS-5 Kap. B 5 Indexieren und B 12 Wissensmanagement). Eine gute Einführung bietet auch Lit. 02.

#### D 1.2.4 Searcher

Gibt der Nutzer eine Suchanfrage ein, so wird der Index abgefragt. Das Anfragemodul (*searcher* oder *query module*) setzt die eingegebene Suchanfrage in eine weiterverarbeitbare Form um. Dabei werden beispielsweise besondere Befehle und Operatoren so aufgelöst, dass sie entsprechend im Index abgefragt werden können. Die Ergebnisse werden aus dem Index gelesen und für den Suchenden aufbereitet.

### D 1.3 Ranking

Kernkomponente jeder Suchmaschine ist das Ranking, welches auf die gefundene Treffermenge angewendet wird. Vor allem bedingt durch das Nutzerverhalten im Web, welches durch wenig elaboreierte Anfragen und die starke Fokussierung auf die ersten Trefferplätze der Ergebnisliste gekennzeichnet ist (Lit. 24, Lit. 25, Lit. 26, Lit. 27), ist ein Ranking nötig, welches sich vor allem auf die Präzision der Suchergebnisse konzentriert. Die Vollständigkeit der Treffermenge (Recall) steht demgegenüber im Hintergrund.

#### D 1.3.1 Rankingfaktoren

Auch wenn die Rankingverfahren der Suchmaschinen nicht offengelegt werden und es sich, wenn man jeden einzelnen Faktor berücksichtigt, um ein Zusammenspiel von hunderten von Faktoren handelt, so haben sich doch vier Bereiche herausgebildet, die für das Ranking der Ergebnisse bestimmend sind.

Mittels *textspezifischer Faktoren* wird abgeglichen, welche Wörter der Suchanfrage in den zu durchsuchenden Dokumenten vorkommen und daher in die Treffermenge mit aufgenommen werden sollen. Das Vorkommen der Suchbegriffe kann sich auch auf im Dokument vorkommende Varianten der Suchbegriffe oder Wörter aus auf das Zieldokument verweisenden Dokumenten (aus den Ankertexten) beziehen (Lit. 28, S. 70). Neben dem Vorkommen der Suchbegriffe wird mittels Textstatistik ein Vorkommen an exponierter Stelle (z. B. Überschrift, Beginn des Dokuments, besondere Hervorhebung) höher gewichtet.

Zweiter bestimmender Bereich im Ranking ist die *Messung der Popularität* von Dokumenten, meist durch deren Verlinkung (Beispiele für Algorithmen: *PageRank*, Lit. 29), *HITS* (Lit. 30), aber auch durch das Klickverhalten der Nutzer (Erfassung über Logfiles, Toolbars).

Dritter Rankingbereich ist die *Aktualität*. Je nach Zweck der Anfrage kann es sinnvoll sein, entweder besonders aktuelle Dokumente oder statische, dafür populäre Dokumente anzuzeigen. Da linktopologische Algorithmen tendenziell ältere Dokumente bevorzugen, wird Aktualität auch als Ausgleichsfaktor herangezogen. In der Praxis sind meist durchmischte Trefferlisten zu finden, in die einige besonders aktuelle Dokumente eingestreut werden.

*Lokalität* berücksichtigt den Standort des Nutzers. Ein einfaches Beispiel ist die Bevorzugung von Dokumenten aus Deutschland, wenn der Standort des Nutzers dort identifiziert wird. Zuneh-

mend wird aber auch eine genaue Standortbestimmung des Nutzers vorgenommen; vor allem auch in Hinblick auf die Suche von mobilen Endgeräten aus.

### D 1.3.2 Personalisierung der Suchergebnisse

Die bislang letzte große Entwicklung im Ranking ist die Personalisierung der Suchergebnisse (Lit. 31). Sie kann in den Bereich der nutzungsstatistischen Verfahren eingeordnet werden: Personalisierung aufgrund des Verhaltens eines einzigen Nutzers erfolgt anhand impliziter Information aus seinem Surfverhalten; explizite Daten können zusätzlich über Bewertungen (entweder durch den Nutzer selbst oder durch seine Kontakte in einem sozialen Netzwerk) gewonnen werden.

Die Personalisierung der Suchergebnisse wird vielfach kritisch gesehen. Sie führe dazu, dass die Ergebnisse der Suchmaschinen vor allem die eigene Meinung bestätigen würden und andere Meinungen unterdrückt würden. Außerdem wären Entdeckungen, wie sie in nicht-personalisierten Suchergebnissen möglich gewesen wären, nun nicht mehr möglich, oder doch zumindest unwahrscheinlicher. Die Kritik von Eli Pariser in seinem Buch *The Filter Bubble* (Lit. 32) berücksichtigt, dass Nutzer schon immer Medien nach ihrem Geschmack ausgewählt hätten, allerdings würde durch die Personalisierung durch Suchmaschinen (und Soziale Netzwerke) eine neue Dimension erreicht: So wäre die Filterblase erstens dadurch gekennzeichnet, dass sie auf jeden Nutzer individuell angepasst sei und jeder Nutzer eben unterschiedliche Ergebnisse zu sehen bekäme, zweitens dadurch, dass sie für den Nutzer unsichtbar bleibe, und drittens, dass sich Nutzer nicht für oder gegen die Personalisierung ihrer Ergebnisse entscheiden könnten, sondern die Verfahren ohne Rückfrage von den Suchmaschinen angewendet werden würden.

Unstrittig ist allerdings, dass die Personalisierung von Suchergebnissen zu einer weit besseren Ergebnisqualität führen kann, da eben gerade die individuelle Anpassung auf den einzelnen Nutzer viele Ergebnisse, die für die Masse relevant sein können, für den einzelnen Nutzer jedoch nicht, ausgefiltert werden können. Die Debatte hat sich hier allerdings verlagert von einer Kritik an den Suchmaschinen dafür, dass sie auf die Masse ausgerichtete, oberflächliche Ergebnisse anzeigen würden hin zu einer Kritik an der (zu starken) Anpassung an den einzelnen Nutzer.

### D 1.3.3 Zur Geschichte der Suchmaschinen und ihres Rankings

Anhand der Rankingfaktoren und der Reihenfolge der vorangegangenen Darstellung lässt sich auch kurz die Geschichte des Suchmaschinenrankings beschreiben. Frühe Suchmaschinen wie Excite, Lycos und AltaVista waren noch stark an konventionellen Information-Retrieval-Systemen orientiert und setzten demnach das dort übliche textstatistische Ranking ein. Unter der Annahme, dass alle Dokumente, die in einen Datenbestand aufgenommen werden, potentiell von gleicher Qualität sind, kann ein solches Ranking (vor allem verbunden mit der Annahme, dass Nutzer bereit und in der Lage sind, treffsichere Suchanfragen zu formulieren) zu guten Ergebnissen führen. Im Kontext des Web, wo die Dokumente von einer Vielzahl von Autoren mit unterschiedlichsten Motiven erstellt werden, muss jedoch zusätzlich die Zuverlässigkeit der Dokumente bestimmt werden. Frühe Suchmaschinen litten unter einer Vielzahl von sog. Spam-Dokumenten, also nicht relevanten Dokumenten, die erstellt wurden, um die Suchmaschinen zu täuschen. So listeten etwa zahlreiche Suchmaschinen zur Suchanfrage *white house* die Website *whitehouse.com*, ein Pornoangebot. Erst die linktopologischen Verfahren (ab etwa 1998) (hier allen voran: Google) konnten dieses Problem lösen.

Dass auch Aktualität eine entscheidende Rolle spielt, wurde im Suchmaschinenbereich klar, als Nutzer nach den Terroranschlägen des 11. September 2001 nach aktuellen Nachrichten zum Thema suchten, die Suchmaschinen allerdings – wenn sie denn überhaupt erreichbar waren – vorwiegend ältere Dokumente anzeigten (Lit. 33), die von den linktopologischen Verfahren tendenzi-

ell bevorzugt werden (Lit. 28, S. 123ff). So wurde Aktualität als Rankingfaktor mit eingeführt; daneben wurden spezielle Kollektionen für aktualitätskritische Inhalte (wie Nachrichten, Blog-Artikel, usw.) aufgebaut.

Während Suchmaschinen schon früh ihre Rankingverfahren an das benutzte Länderinterface angepasst hatten und für die aus unterschiedlichen Ländern gestellte gleiche Suchanfrage die Ergebnisse in unterschiedlicher Reihung ausgaben, erfolgt seit etwa 2008 eine Anpassung der Ergebnisse auf den konkreten Standort des Nutzers.

### D 1.3.4 Suchmaschinenoptimierung

Unter Suchmaschinenoptimierung (*search engine optimization*; *SEO*, Lit. 34, Lit. 35) versteht man alle Maßnahmen, nicht etwa die Qualität der Suchmaschine zu verbessern, sondern die dazu geeignet sind, die Position von Webseiten im Ranking der Suchmaschinen zu verbessern. Die Maßnahmen reichen von einfachen technischen Maßnahmen, die dabei helfen, die Dokumente überhaupt für Suchmaschinen indexierbar zu machen bis hin zu komplexen Manipulationen der Verlinkungsstruktur der auf die zu optimierenden Dokumente verweisenden Seiten. Suchmaschinenoptimierung steht im Spannungsfeld zwischen einer Hilfeleistung auch für die Suchmaschinen und deren Manipulation. Man spricht bei einer Optimierung nach den Regeln der Suchmaschinenbetreiber von *white hat*, und im Gegensatz dazu von *black hat*, wenn es sich um gezielte Manipulationen handelt, die die Nutzungsbedingungen der Suchmaschinenbetreiber unterlaufen.

Suchmaschinenoptimierung ist ein Bestandteil des Suchmaschinenmarketings (*search engine marketing*, *SEM*; siehe Lit. 36; vgl. C 4 Online-Marketing); beide Begriffe werden jedoch häufig fälschlicherweise gleichgesetzt. Suchmaschinenmarketing bezeichnet alle Marketingmaßnahmen, die mittels Suchmaschinen durchgeführt werden, zu nennen ist hier neben der Suchmaschinenoptimierung die Platzierung von Textanzeigen auf Suchergebnisseiten (*search engine advertising*, *SEA*).

The screenshot shows a Google search for 'geld'. The search bar at the top contains 'geld' and a search icon. Below the search bar, it says 'Suche' and 'Ungefähr 315 000 000 Ergebnisse (0,21 Sekunden)'. The results are categorized on the left: Web, Bilder, Maps, Videos, News, Shopping, Mehr, Berlin Standort ändern, Web Seiten auf Deutsch, Seiten aus Deutschland, Übersetzte Seiten, and Mehr Optionen.

The main content area is divided into several sections:

- Anzeige zu geld**: A sponsored result from GELD.de, titled 'GELD.de Ihr Finanzexperte | GELD.de'. It includes a star rating and promotional text about top products and financing offers.
- Geld - Wikipedia**: A snippet from Wikipedia defining 'Geld' and listing related terms like 'Falschgeld', 'Geldfunktion', and 'Geldschöpfung'.
- Bilder zu geld**: A section showing image thumbnails of various banknotes and coins.
- Geld Zitate und Sprüche**: A snippet from zitrate.net/html with the quote: 'Die besten Dinge im Leben sind nicht die, die man für Geld bekommt.'
- Geld.de**: A snippet from www.geld.de/ comparing insurance options.
- Anzeigen**: A column of additional sponsored results on the right, including 'Geld leihen in 2 Minuten', 'Online Geld Verdienen', 'Ria Geldtransfer', 'Verdienen Sie mit Forex', and 'Kredit bis 9.000€'.

On the right side, there is a legend with three boxes: 'Anzeigen', 'Universal Search-Ergebnisse', and 'Organische Ergebnisse'. Arrows point from these boxes to the corresponding sections in the search results: 'Anzeigen' points to the GELD.de ad and the 'Anzeigen' column; 'Universal Search-Ergebnisse' points to the Wikipedia and 'Geld Zitate und Sprüche' results; 'Organische Ergebnisse' points to the 'Geld.de' result.

Abb. 2: Universal-Search-Ergebnisseite (Beispiel Google)



In der Trefferpräsentation von Suchmaschinen ist zwischen dem sichtbaren und dem unsichtbaren Bereich zu unterscheiden (Lit. 38). Ersterer bezeichnet den ohne Scrollen auf einem Bildschirm sichtbaren Bereich der Trefferseite, wobei dieser je nach Bildschirmauflösung (und eingestellter Fenstergröße) unterschiedlich groß sein kann. Als unsichtbarer Bereich wird derjenige Teil einer Trefferseite bezeichnet, der erst durch Scrollen sichtbar wird. Studien haben gezeigt, dass ein großer Teil der Nutzer von Web-Suchmaschinen auf den Ergebnisseiten nicht scrollt, und dass der wesentliche Teil der Klicks auf Ergebnisse im sichtbaren Bereich entfällt (Lit. 40).

### D 1.5 Bedeutung der Suchmaschinen für die Informationsbeschaffung, Problembereiche

Suchmaschinen sind heute die vorherrschenden Informationssysteme, was sich in ihrer enormen Popularität ausdrückt. Dies hat zur Folge, dass sie sich auf der einen Seite an das Verhalten ihrer Nutzer anpassen, dieses auf der anderen Seite aber auch prägen. Sie setzen damit Trends und Nutzer erwarten von anderen Information-Retrieval-Systemen eine vergleichbar einfache Bedienung und eine ähnliche Trefferaufbereitung und -qualität. Es zeigt sich allerdings auch, dass Suchmaschinen universell eingesetzt werden und Nutzer selten vor einer Recherche fragen, ob die Suchmaschine überhaupt das am besten geeignete Informationssystem zur Befriedigung des aktuellen Informationsbedürfnisses ist.

Angesichts der enormen Bedeutung der Suchmaschinen ist es verwunderlich, wie wenig über die Qualität dieser Suchwerkzeuge und die Auswirkungen der starken Fokussierung der Nutzer auf Suchmaschinen für Recherchen aller Art bekannt ist. Wenn Recherchen sowohl im privaten als auch im professionellen Bereich vor allem über Suchmaschinen durchgeführt werden, so ist zu fragen, welche Treffer sie anzeigen, welche ausgewählt werden und welchen Einfluss die Suchmaschinen mittels ihrer Rankingalgorithmen auf die Informationsauswahl und letztlich die Wissensaneignung innerhalb der Gesellschaft haben.

In dieser Hinsicht sind vor allem vier Problembereiche zu sehen: Externer Einfluss auf die Suchergebnisse durch Suchmaschinenoptimierung, Werbetreffer, Verzerrungen und Monopolisierung des Suchmaschinenmarkts.

Das Themenfeld *Suchmaschinenoptimierung* wurde oben hinsichtlich seiner Vorgehensweisen und seiner Möglichkeiten beschrieben. Im positiven Sinne leistet Suchmaschinenoptimierung einen Beitrag zur Erschließung der Inhalte durch Suchmaschinen; auf der anderen Seite kann sie eine Manipulation der Ergebnislisten darstellen, die so weit gehen kann, dass Spam-Dokumente hoch platziert angezeigt werden. Während die Suchmaschinenoptimierung mittlerweile eine etablierte Branche ist und kaum ein kommerzielles Unternehmen ohne ihre Dienste auskommt, ist der Einfluss der Suchmaschinenoptimierung auf die Informationsrecherche noch weitgehend unerforscht. Vor allem bei informationsorientierten Suchanfragen (im Gegensatz zur Suche nach Produkten und Dienstleistungen) besteht hier großer Bedarf.

Suchmaschinen sind nicht zuletzt auch Instrumente zur Generierung unternehmerischer Gewinne. Werbetreffer generieren den weit überwiegenden Umsatz der Suchmaschinenbetreiber, andere Erlöse spielen nur eine untergeordnete Rolle (Lit. 41). Werbung in Suchmaschinen ist so erfolgreich, weil sie kontextbasiert ist, d. h. auf Suchanfragen hin angezeigt wird. Ein Nutzer, der eine Suchanfrage eingibt, offenbart bereits ein Interesse. John Batelle spricht in diesem Zusammenhang von der *database of intentions* (Lit. 42), über die Suchmaschinen verfügen. Nimmt man nun die durch den Abgleich der Suchanfrage mit der Werbung erreichte Relevanz zusammen mit der Darstellung der Werbetreffer, die den organischen Ergebnissen ähnelt, sowie die Positionierung der Werbetreffer direkt oberhalb der organischen Ergebnisse zusammen, so liegt die Vermutung nahe, dass zumindest ein Teil der Nutzer nicht in der Lage ist, die Werbetreffer von organischen Treffern zu unterscheiden. Bisherige Studien (welche jedoch nur beschränkte Teilnehmerzahlen aufweisen) deuten auf diesen Umstand hin (Lit. 43, Lit. 44, Lit. 45).

*Verzerrungen* (Bias) in Suchmaschinen ergeben sich auf verschiedenen Ebenen (Lit. 46, Lit. 47). So wird den bekannten Suchmaschinen unter anderem vorgeworfen, sie würden ihre eigenen Angebote in den organischen Trefferlisten (Lit. 48, Lit. 49) und bei *den Universal-Search*-Ergebnissen (Lit. 38) bevorzugen. Aus diesem Grund ist eine Debatte über die sog. Such-Neutralität entbrannt; im Kern geht es um die Verantwortung der Suchmaschinenbetreiber für eine klare Kennzeichnung ihrer Interessenskonflikte auf den Suchergebnisseiten.

Als letzter wichtiger Problembereich ist die *Aufteilung des Suchmaschinenmarkts* zu nennen. Wie oben beschrieben, besteht in den europäischen Ländern ein Quasi-Monopol durch Google, welches nicht nur zu einer mangelnden Vielfalt führt, sondern dieser Suchmaschine auch die Erstellung von Nutzerprofilen erlaubt, die bei einer Verteilung der Anfragen auf mehrere Suchmaschinen nicht möglich ist. Eine wesentliche Schwierigkeit für neue Suchmaschinen stellt neben der Marktdominanz von Google (und seiner zugegebenermaßen guten Qualität) der finanzielle und technische Aufwand für die Erstellung und Pflege eines Web-Index dar. Würde man die Dominanz einer einzigen Suchmaschine aufbrechen wollen, so wäre die Bereitstellung eines Web-Index, der für alle Marktteilnehmer zu gleichen und fairen Konditionen abfragbar ist, als erster Schritt anzusehen.

## D 1.6 Ausblick

Ein Blick in die Zukunft ist immer schwierig – nichtsdestotrotz können bereits heute Bereiche genannt werden, die zukünftige Suchmaschinen prägen werden:

Suchmaschinen entfernen sich schon heute von der üblichen Ausgabe einer Trefferliste auf eine gestellte Suchanfrage hin. Wir werden in Zukunft eine Abwendung von den Dokumenten hin zu Antworten und von der Suchmaschine aus Dokumenten zusammengestellten Informationen erleben. Die Trefferseite (SERP) wird zunehmend zum Ergebnis selbst.

Die zweite Entwicklung betrifft die Eingabe von Suchanfragen. Während heute die meisten Suchen noch durch eine Suchanfrage ausgelöst werden, wird in Zukunft die weit überwiegende Zahl von Suchanfragen aufgrund von Kontextinformationen automatisch generiert werden. Wir können heute schon sehen, dass auf Portalen explizite Suchen von Navigationsklicks nicht mehr unterschieden werden können; im Fall eines Klicks wird oft eine Suchanfrage ausgelöst, ohne dass der Nutzer dies merken würde. Betrachtet man nun vor allem mobile Endgeräte, so können kontextbasierte Suchanfragen beispielsweise aus einer Kombination des Standorts des Nutzers, seiner Kontakte, seiner aus vergangenen Aktionen bekannten Präferenzen und der aktuellen Zeit generiert werden. So kann einem Nutzer beispielsweise am Nachmittag ein Café empfohlen werden, das seinen Präferenzen entspricht und in dem er aktuell Freunde treffen kann, die sich gerade dort aufhalten.

In diesem Beispiel werden Suchanfragen an verschiedene Systeme geschickt und die Ergebnisse kombiniert, allerdings ohne dass der Nutzer explizit eine Suchanfrage eingibt. Wir sehen hier das Verschmelzen von Suche und Empfehlungssystemen.

Damit wird Suche allgegenwärtig, auch wenn sich unser Verständnis von Suche damit verändert. Suche wird nicht nur bestehen bleiben, sondern weiterhin an Bedeutung gewinnen. Was sich ändern wird, ist die Sichtbarkeit der Suche. Diese wird zurückgehen.

## Literatur

- 01 Büttcher, S.; Clarke, C.L.A.: *Information Retrieval: Implementing and Evaluating Search Engines*. MIT Press, Cambridge, Mass, 2010
- 02 Croft, W.B.; Metzler, D.; Strohman, T.: *Search Engines: Information retrieval in practice*. Pearson, Boston, MA, 2010
- 03 Baeza-Yates, R.; Ribeiro-Neto, B.: *Modern Information Retrieval: The Concepts And Technology Behind*

- Search. Addison Wesley, Harlow, 2011
- 04 Eimeren, B.V.; Frees, B.: Drei von vier Deutschen im Netz – ein Ende des digitalen Grabens in Sicht? *Media Perspektiven*, 334-349, 2011
  - 05 Griesbaum, J.; Bekavac, B.; Rittberger, M.: Typologie der Suchmaschine im Internet. Lewandowski, D. (ed.) *Handbuch Internet-Suchmaschinen*, 18-52. Akademische Verlagsgesellschaft Aka GmbH, Heidelberg, 2009
  - 06 Hamdorf, K.: Jenseits von Google – Erschließung und Recherche von Internet-Angeboten durch Webkataloge. *Information Wissenschaft und Praxis*. 55, 22-224, 2004
  - 07 Peters, I.: Folksonomies und Kollaborative Informationsdienste: Eine Alternative zur Websuche? Lewandowski, D. (ed.) *Handbuch Internet-Suchmaschinen 2: Neue Entwicklungen in der Web-Suche*, 29–53. Akademische Verlagsanstalt AKA, Heidelberg, 2011
  - 08 Gazan, R.: Advances in Information Science: Social Q & A. *Journal of the American Society for Information Science and Technology*, 62, 2301-2312, 2011
  - 09 Thomas, P.: To What Problem Is Distributed Information Retrieval the Solution? *Journal of the American Society for Information Science & Technology*. 63, 1471-1476, 2012
  - 10 Lewandowski, D.: Spezialsuchmaschinen. Lewandowski, D. (ed.) *Handbuch Internet-Suchmaschinen*, 53–69. AKA, Heidelberg, 2009
  - 11 Broder, A.: A taxonomy of web search. *ACM Sigir forum*, 36, 3-10, 2002
  - 12 Lewandowski, D.; Drechsler, J.; Mach, S.V.: Deriving Query Intents From Web Search Engine Queries. *Journal of the American Society for Information Science and Technology*, 63, 1773-1788, 2012
  - 13 Press Release: comScore Reports Global Search Market Growth of 46 Percent in 2009 ([http://comscore.com/Press\\_Events/Press\\_Releases/2010/1/Global\\_Search\\_Market\\_Grows\\_46\\_Percent\\_in\\_2009](http://comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_Percent_in_2009))
  - 14 Sterling, G.: August Search Share: Bing Hits 'All Time High' (<http://searchengineland.com/august-search-share-bing-hits-all-time-high-133021>)
  - 15 Maaß, C.; Skusa, A.; Heß, A.: Der Markt für Internet-Suchmaschinen. *Handbuch Internet-Suchmaschinen*, 3-17, 2009
  - 16 Webhits: Webhits Web-Barometer (<http://www.webhits.de/deutsch/index.shtml?webstats.html>)
  - 17 Risvik, K.M.; Michelsen, R.: Search engines and web dynamics. *Computer Networks*, 39, 289-302, 2002
  - 18 Vaidhyanathan, S.: *The Googlization of Everything (and why we should worry)*. University of California Press, Berkeley, CA, 2011
  - 19 Broder, A.; Kumar, R.; Maghoul, F.; Raghavan, P.; Rajagopalan, S.; Stata, R.; Tomkins, A.; Wiener, J.: Graph structure in the web. *Computer networks*, 33, 309-320, 2000
  - 20 Sherman, C.; Price, G.: *The Invisible Web: Finding Hidden Internet Resources Search Engines Can't See*. Cyberage Books, 2001
  - 21 Bergman, M.K.: The deep Web: Surfacing hidden value. *Journal of Electronic Publishing*, 7, 1-17, 2001
  - 22 Lewandowski, D.; Mayr, P.: Exploring the academic invisible web. *Library Hi Tech*. 24, 529-539, 2006
  - 23 Madhavan, J.; Ko, D.; Kot Kot, L.; Ganapathy, V.; Rasmussen, A.; Halevy, A.: Google's Deep-Web Crawl. *Conference on Very Large Data Bases*, 1241–1252. VLDB Endowment, Auckland. 2008
  - 24 Spink, A.; Jansen, B.J.: *Web search: public searching on the Web*. Kluwer Academic Publishers, Dordrecht, 2004
  - 25 Schmidt-Mänz, N.: *Untersuchung des Suchverhaltens im Web: Interaktion von Internetnutzern mit Suchmaschinen*. Verlag Dr. Kovac, 2007
  - 26 Lorigo, L.; Haridasan, M.; Brynjarsdóttir, H.; Xia, L.; Joachims, T.; Gay, G.; Granka, L.; Pellacini, F.; Pan, B.: Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, 59, 1041-1052, 2008
  - 27 Pan, B.; Hembrooke, H.; Joachims, T.; Lorigo, L.; Gay, G.; Granka, L.: In Google we trust: users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, 12, 801-823, 2007
  - 28 Lewandowski, D.: *Web Information Retrieval: Technologien zur Informationssuche im Internet*. Deutsche Gesellschaft f. Informationswissenschaft u. Informationspraxis, Frankfurt am Main, 2005
  - 29 Page, L.; Brin, S.; Motwani, R.; Winograd, T.: *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab, 1999

- 30 Kleinberg, J.: Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46, 604-632, 1999
- 31 Riemer, K.; Brüggemann, F.: Personalisierung der Internetsuche – Lösungstechniken und Marktüberblick. Lewandowski, D. (ed.) *Handbuch Internet-Suchmaschinen*, 148–171. Akademische Verlagsgesellschaft Aka, Heidelberg, 2009
- 32 Pariser, E.: *The Filter Bubble: What The Internet Is Hiding From You*. Viking, London, 2011
- 33 Wiggins, R.W.: The Effects of September 11 on the Leading Search Engine. *First Monday*, 6, 20-21, 2001
- 34 Fischer, M.: *Website Boosting 2.0: Suchmaschinen-Optimierung, Usability, Online-Marketing*, 2009
- 35 Erlhofer, S.: *Suchmaschinen-Optimierung für Webentwickler*. Galileo Press, Bonn, 2007
- 36 Schultz, C.D.: *Suchmaschinenmarketing*. Lewandowski, D. (ed.) *Handbuch Internet-Suchmaschinen*, 70-98. Akademische Verlagsgesellschaft Aka GmbH, Heidelberg, 2009
- 37 Quirnbach, S.: *Universal Search – Kontextuelle Einbindung von unterschiedlicher Quellen und Auswirkungen auf das User Interface*. Lewandowski, D. (ed.) *Handbuch Internet-Suchmaschinen*, 220-248. Akademische Verlagsgesellschaft Aka GmbH, Heidelberg, 2009
- 38 Höchstätter, N.; Lewandowski, D.: What users see – Structures in search engine results pages. *Information Sciences*, 179, 1796-1812, 2009
- 39 Jansen, J.: *Understanding sponsored search: Core elements of keyword advertising*. Cambridge University Press, New York, 2011
- 40 Höchstätter, N.; Koch, M.: Standard parameters for searching behaviour in search engines and their empirical evaluation. *Journal of Information Science*, 35, 45-65, 2009
- 41 Google Annual Report, 2011
- 42 Batelle, J.: *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. London: Portfolio. Brealey, London [u. a.], 2005
- 43 Fallows, D.: *Search engine users: Internet searchers are confident, satisfied and trusting – but they are also unaware and naive*. Pew Internet & American Life Project, Washington, DC, 2005
- 44 Marable, L.: *False oracles: consumer reaction to learning the truth about how search engines work*. research report, 30, 1-66, 2003
- 45 Bundesverband Digitale Wirtschaft: *Nutzerverhalten auf Google-Suchergebnisseiten: Eine Eyetracking-Studie im Auftrag des Arbeitskreises Suchmaschinen-Marketing des Bundesverbandes Digitale Wirtschaft (BVDW) e.V.*, 2009
- 46 Weber, K.: *Search Engine Bias*. Lewandowski, D. (ed.) *Handbuch Internet-Suchmaschinen 2: Neue Entwicklungen in der Web-Suche*, 265–285. Akademische Verlagsanstalt AKA, Heidelberg, 2011
- 47 Röhle, T.: *Der Google-Komplex: Über Macht im Zeitalter des Internets*. Transcript, Bielefeld, 2010
- 48 Edelman, B.: *Hard-Coding Bias in Google „Algorithmic“ Search Results* (<http://www.benedelman.org/hardcoding/>)
- 49 Edelman, B.; Lockwood, B.: *Measuring bias in „organic“ web search* (<http://www.benedelman.org/searchbias/>)