

---

## Glossar

**Above the fold** dt. über dem Knick; auch „sichtbarer Bereich“: Der Teil der Suchergebnisseite, der sofort zu sehen ist, ohne dass gescrollt werden muss. Die Größe dieses Teils hängt von der Bildschirm- bzw. Fenstergröße ab. Der Begriff leitet sich von einer gefalteten Zeitung ab.

**Absprungrate** Anteil der Nutzer, die nach dem Aufruf eines Dokuments dieses schnell wieder verlassen (und im Fall der Herkunft von einer Suchmaschine auf die Suchergebnisseite zurückkehren).

**Adresszeile** Eingabezeile im Browser, in die eine URL eingegeben wird, um sie direkt aufzurufen.

**AdSense** Anzeigenprogramm von Google, bei dem Websitebetreiber Anzeigen auf ihren eigenen Dokumenten schalten können. Die Anzeigen werden passend zum Text des Dokuments generiert.

**AdWords** Anzeigenprogramm von Google, bei dem Websitebetreiber Textanzeigen erstellen, die auf den Suchergebnisseiten der Suchmaschine angezeigt werden. Die Buchung erfolgt auf Suchwörter; der Preis wird in einem Bieterverfahren zwischen den Werbetreibenden festgelegt. Teilweise wird AdWords auch allgemein als Bezeichnung für Textanzeigen in Suchmaschinen verwendet.

**Affiliate-Marketing** Marketing über Vertriebspartner, die Nutzer auf die Seite des Anbieters lenken und dafür eine Provision bekommen.

**Aggregatoren** Anbieter, die Informationen aus unterschiedlichen Quellen zusammenführen.

**Aktualisierungsdatum** Datum der letzten inhaltlichen Aktualisierung eines Dokuments.

**Algorithmus** Eine eindeutige Folge von Handlungsanweisungen, welche nacheinander abgearbeitet werden. Im Kontext von Suchmaschinen ist meist der Algorithmus bzw. sind die Algorithmen gemeint, nach denen die Suchergebnisse sortiert werden.

**Anfrageinterpretation** *engl. query understanding*: Hinzufügen von Kontextinformationen zu gestellten Suchanfragen, beispielsweise dem Standort des Nutzers, um

Suchanfragen gezielter beantworten zu können. Man unterscheidet zwischen expliziter und impliziter Anfrageinterpretation, je nachdem, ob die hinzugefügten Informationen dem Nutzer mitgeteilt werden oder nicht.

**Anfragetyp** *engl. query intent*: Unterscheidung von Suchanfragen nach dem Ziel bzw. dem hinter der Suchanfrage stehenden Informationsbedürfnis. Die für die Websuche bedeutendste Einteilung unterscheidet zwischen informationsorientierten (*information*), navigationsorientierten (*navigational*) und transaktionsorientierten (*transactional*) Suchanfragen.

**Ankertext** *engl. anchor text*: Text innerhalb eines Dokuments, der mit einem anderen Dokument verlinkt ist. Ankertexte werden von Suchmaschinen zur ergänzenden Beschreibung des Dokuments, auf das verlinkt wird, verwendet.

**Anzeige** s. Textanzeige

**App** Anwendungssoftware; als Apps werden insbesondere Anwendungen, die für mobile Endgeräte verfügbar sind, bezeichnet.

**Application Programming Interface (API)** dt. Programmierschnittstelle: Schnittstelle, die einen automatisierten Zugriff auf eine Software erlaubt; im Fall der Suchmaschinen die automatische Abfrage von Suchergebnissen.

**Autocomplete** s. Suchvorschlag

**Autokorrektur** Automatische Korrektur von (vorseiten der Suchmaschine angenommen) Fehleingaben, v. a. Tippfehler.

**Befehl** Kommando, das zur Qualifizierung einer Suchanfrage eingesetzt wird.

**Below the fold** dt. unter dem Knick; auch „unsichtbarer Bereich“: Der Bereich einer Suchergebnisseite, der erst durch Scrollen erreichbar ist.

**Beziehungsgeflecht der Suchmaschinen** *engl. search engine relationship chart*: Grafische Darstellung der Lieferungen von Suchergebnissen und Textanzeigen zwischen verschiedenen Suchmaschinen.

**Bias** dt. Verzerrung: Im Kontext der Suchmaschinen wird darunter die Abweichung der tatsächlichen Suchergebnisse von einem angenommenen idealen Ergebnis erwartet. Hervorgerufen wird der Bias durch technische Probleme in der Indexierung, durch Faktoren des Rankings, durch gezielte Bevorzugung bestimmter Angebote durch die Suchmaschinenbetreiber und durch externe Manipulationen der Suchergebnisse (s. Suchmaschinenoptimierung).

**Black Hat** Im Kontext der Suchmaschinenoptimierung Methoden, die von den Suchmaschinen in ihren Nutzungsbedingungen ausgeschlossen werden.

**Black List** Liste von Dokumenten bzw. Websites, die aus dem Index einer Suchmaschine ausgeschlossen werden.

**Bookmark** dt. Lesezeichen: Link, der im Browser für einen schnellen Zugriff abgelegt wird. Lesezeichen erlauben das schnelle Wiederfinden und Aufrufen vorseiten im World Wide Web.

**Boolesche Suchanfrage** Anfrage, die mithilfe von Booleschen Operatoren formuliert wird

**Boolesche Operatoren** *engl. Boolean operators*: Werden verwendet, um logische Beziehungen von Elementen in Suchanfragen auszudrücken. Die Booleschen Operatoren

sind AND, OR und NOT; weitere auf den Booleschen Operatoren basierende Operatoren wurden entwickelt.

**Boost** *dt. Schub*: Bestimmten Dokumenten einen *boost* zu geben, bedeutet, ihnen im Ranking einen Schub nach oben in den Trefferlisten zu geben.

**Bow-Tie-Modell** s. Fliegen-Modell

**Browser** Software, mit deren Hilfe Dokumente im World Wide Web angezeigt werden können.

**Cache** Zwischenspeicher einer Suchmaschine, in der die zuletzt vom Crawler besuchte Version des Dokuments abgelegt ist.

**Click Fraud** s. Klickbetrug

**Cloaking** Ausspielen unterschiedlicher Inhalte auf derselben URLs je nach anfragendem Client. So können beispielsweise für menschliche Nutzer und für Suchmaschinen unterschiedliche Inhalte ausgespielt werden, um Suchmaschinen über die wahren Inhalte von Dokumenten zu täuschen.

**Cluster** *dt. Klumpen*: Automatisch zusammengestellte Menge von als ähnlich erkannten Objekten.

**Content Acquisition** Die Aufnahme von Inhalten in den Datenbestand einer Suchmaschine. Diese geschieht vor allem über das Crawling. Allerdings können Inhalte auch auf andere Weise in den Datenbestand einer Suchmaschine gelangen, etwa durch das Auswerten von Feeds und den Zugriff auf strukturierte Daten aus Datenbanken.

**Content-Management-System** Software zur Erstellung und Verwaltung von Inhalten auf Websites. Die dort eingegebenen Inhalte werden vom System automatisch in die Struktur und das Layout der Website eingesetzt.

**Country Bias** Verzerrung in der Abdeckung der Inhalte aus verschiedenen Ländern, d. h. die Inhalte aus bestimmten Ländern werden im Index einer Suchmaschine vollständiger abgedeckt als die Inhalte aus anderen Ländern.

**Crawler** System zum Auffinden von Inhalten im World Wide Web durch das Verfolgen von Links in Dokumenten.

**Crawling** Crawling bezeichnet das Auffinden von Dokumenten im Web durch das Verfolgen von Links auf bereits bekannten Seiten

**Crawling-Frequenz** *engl. crawl frequency*: Häufigkeit der Wiederholung des Crawling eines Dokuments.

**Dark Web** Nur über spezielle Tools zugänglicher Bereich des Internet, der vor allem für den Verkauf von illegalen Produkten und die Distribution von illegalen und urheberrechtlich geschützten Inhalten bekannt ist.

**Data Center** Rechenzentrum einer Suchmaschine. Websuchmaschinen sind auf verschiedene Data Centers verteilt, um eine effiziente Bearbeitung des Crawlings und der Suchanfragen zu gewährleisten.

**Datenbasis der Suchmaschine** *engl. local store*: Die Kopie der Inhalte des Web, die von einer Suchmaschine im Vorgang des Crawlings erstellt wird. Im Idealfall ist diese Kopie ein vollständiges und aktuelles Abbild des World Wide Web. Man spricht hier teilweise auch vom Index.

- Datenerhebung** In nutzungsstatistischen Verfahren können Daten über das Nutzerverhalten entweder über vom Nutzer explizit gemachte Bewertungen oder über die Auswertung von Klicks (implizit) erhoben werden.
- Dauerbrenner** Suchanfrage, die für Nutzer dauerhaft von Interesse ist und deren Suchvolumen daher keinen größeren Schwankungen unterliegt.
- Dedicated Searcher** Modell des Suchverhaltens, das in den klassischen TREC-Studien verwendet wird. Es wird von einem Nutzer ausgegangen, der zum einen alle für sein Thema relevanten Ergebnisse erhalten möchte und zum anderen bereit ist, große Treffermengen durchzusehen.
- Deep Web** Alle über das Web zugänglichen Inhalte, die von den allgemeinen Suchmaschinen nicht erschlossen werden (können), vor allem die Inhalte von Datenbanken, die über das Web erreichbar sind.
- Dokument** Materielle Einheit eines Trägers dokumentarischer Daten. Im Kontext der Suchmaschinen spielen Textdokumente, aber u. a. auch Bilder und Videos eine Rolle.
- Dokumenttyp** Art des Dokuments, beispielsweise Text, Bild oder Video.
- Domainhoster** Anbieter, der Websites im Auftrag in seinem Rechenzentrum betreibt.
- Dublette** Doppelexemplar; im Kontext der Suche gleicher bzw. sehr ähnlicher Inhalt in unterschiedlichen Dokumenten.
- Duplicate Content** s. Dublette
- Dwell time** s. Verweildauer
- Einfache Suche** Standardsuchformular, das aus nur einem Eingabefeld besteht.
- Eintagsfliege** Suchanfrage, die durch ein plötzlich auftauchendes Interesse der Nutzer kurzzeitig ein hohes Suchvolumen erreicht. Eintagsfliegen werden vor allem durch externe Impulse (beispielsweise Berichte in den Medien) ausgelöst.
- Erstellungsdatum** Datum, an dem ein Dokument erstellt wurde, s.a. → Aktualisierungsdatum.
- Erweiterte Suche** engl. *advanced search (interface)*: Suchformulare mit mehreren Eingabefeldern zur Spezifizierung von Suchanfragen.
- Event** Suchanfrage, die zu regelmäßigen Zeitpunkten ein hohes Suchvolumen erreicht (bspw. Weihnachten, Oktoberfest, Hafengeburtstag).
- Explizite Anfrageinterpretation** s. Anfrageinterpretation
- Explizite Datenerhebung** s. Datenerhebung
- Eyetracking** Verfahren, bei dem die Bewegungen der Pupillen von Probanden mittels Infrarotkameras gemessen werden. Daraus lassen sich typischen Blickverläufe und Blickhäufungen messen.
- F-Muster** engl. *f pattern*: Typischer Blickverlauf auf Suchergebnisseiten, die aus einer Trefferliste bestehen, über der ein Block mit Textanzeigen angezeigt wird. Die Blicke setzen bei den Anzeigen an, dann springen sie auf das erste organische Suchergebnis.
- Fakteninformation(en)** Im Gegensatz zu Textdokumenten können Fakteninformationen als direkte Antworten ausgegeben werden, d. h. es ist nicht erforderlich, den Nutzer von der Suchergebnisseite aus auf ein Dokument weiterzuleiten.
- Faktor** s. Rankingfaktor
- Feeds** Feeds bieten eine Möglichkeit, dem Index einer Suchmaschine Daten in strukturierter Form hinzuzufügen.

**File Transfer Protocol** Protokoll zur Übertragung von Dateien über IP-Netzwerke.

**Flash** Dateiformat für multimediale und interaktive Inhalte. Suchmaschinen können Flash-Inhalte nicht vollständig erfassen.

**Fliegen-Modell** engl. *bow-tie model*. Modell der Verlinkungen im World Wide Web, das zwischen einem stark verlinkten Kernbereich und verschiedenen Außenbereichen unterscheidet. Aus der Struktur ergeben sich Probleme in der vollständigen und gleichmäßigen Erfassung der Web-Dokumente im Crawling.

**Focused Crawling** s. fokussiertes Crawling

**Fokussiertes Crawling** engl. *focused crawling*: Crawling, das sich willentlich auf einen bestimmten Bereich des Web beschränkt.

**FTP** s. File Transfer Protocol

**Global Positioning System** s. GPS

**Goldenes Dreieck** engl. *golden triangle*: Typisches Blickverlaufsmuster, bei dem die ersten Dokumente in einer Liste am häufigsten und am intensivsten betrachtet werden.

**Google Analytics** Software von Google, die Websitebetreibern eine Analyse der Besucherströme erlaubt.

**Google Search Console** Von Google bereitgestellte Software, die bei der Suchmaschinenoptimierung unterstützt (ehemals Google Webmaster Tools).

**GPS** Global Positioning System: mobile Endgeräte können ihren genauen Standort übermitteln, was eine genaue Anpassung der Suchergebnisse auf diesen ermöglicht.

**Heatmap** Darstellungsform für Ergebnisse aus Eyetracking-Untersuchungen.

**HITS** Von Jon Kleinberg entwickeltes linktopologisches Rankingverfahren, das Dokumente u. a. in Hubs und Authorities teilt.

**Horizontale Suchmaschine** Allgemeine Suchmaschine, die den Anspruch hat, die Inhalte des Web in ihrer Breite zu erfassen.

**Hosting** Technische Bereitstellung von Websites.

**HTML (Hypertext Markup Language)** Auszeichnungssprache für Dokumente im World Wide Web.

**HTML-Quelltext** s. Quelltext

**HTML-Seite** Synonym für Webseite; auch Seite, Dokument.

**Hybridsuchmaschine** Suchmaschine, die sowohl Teile des Web (s. fokussiertes Crawling) als auch Inhalte aus Invisible-Web-Datenbanken erfasst und diese gemeinsam durchsuchbar macht.

**Implizite Anfrageinterpretation** s. Anfrageinterpretation

**Implizite Datenerhebung** s. Datenerhebung

**Impuls** Suchanfrage, die durch ein externes Ereignis kurzfristig ein hohes Suchvolumen erreicht, das dann langsam abflaut.

**In-Link** Ein Link, der auf ein bestimmtes Dokument zeigt, wird als In-Link dieses Dokuments bezeichnet.

**Index** Bereits für die Suche aufbereiteter Datenbestand einer Suchmaschine.

**Indexer** Komponente der Suchmaschine, die für die Aufbereitung der Dokumente zuständig ist.

- Information Retrieval** Computergestütztes Suchen bzw. Wiederfinden von Informationen zu einer speziellen Fragestellung.
- Informationsbedarf** Art und Umfang der für eine Problemstellung benötigten Informationen.
- Informationsbedürfnis** Ausdruck dessen, was eine Recherchierender zu wissen wünscht.
- Informationskompetenz** Fähigkeit, kompetent mit Informationen und Informationssystemen umzugehen.
- Informationsobjekt** engl. *information object*: Dokument, unabhängig von einem Dokumenttyp (wie Text oder Bild).
- Informetrische Verteilung** Stark ungleichmäßige Verteilung, bei der wenige Objekte einen Großteil des Volumens auf sich vereinen, während viele Objekte nur ein geringes Volumen aufweisen. Beispiele sind die Verteilung von Wörtern in Dokumenten und die Verteilung von Links im World Wide Web.
- Inverse Dokumenthäufigkeit** engl. *inverted document frequency*: Berechnung der Bedeutung von Wörtern innerhalb von Dokumentkollektionen, womit seltenere Wörter stärker gewichtet werden können.
- Invertierter Index** Ein invertierter Index kehrt die Dokumente um von Texten, die Wörter enthalten zu Wortlisten, die auf Dokumente verweisen.
- Invisible Area** s. Unsichtbarer Bereich.
- Invisible Web** s. Deep Web
- IP-Adresse** Eindeutige Kennung eines Rechners, die benötigt wird, um diesem über das Internet vermittelte Daten zuzuleiten.
- Iteratives Verfahren** Verfahren, das in mehreren Schritten annäherungsweise Werte errechnet, ohne dass am Anfang bereits bekannte Werte stehen.
- Juror** Person, die im Rahmen eines Retrievaltests die Relevanz von Dokumenten bewertet.
- Keyword** Suchwort.
- Keyword stuffing** dt. künstliche Worthäufungen: Häufung von potentiellen Suchwörtern innerhalb eines Dokuments mit dem Ziel, die Suchmaschinen zu täuschen.
- Keywords in Context (KWIC)** Darstellungsform zu Exzerpte aus Dokumenten (beispielsweise in Trefferbeschreibungen), bei der ein Ausschnitt, der die eingegebenen Suchwörter enthält, dargestellt wird.
- Klickbetrug** Vorgehen, bei dem manuell oder automatisch Klicks auf Anzeigen generiert werden, um den Werbetreibenden, der pro Klick bezahlt, zu schädigen.
- Knowledge Graph** Zusammenstellung von Fakten in einem Container auf einer Suchergebnisseite.
- Kollektion** Datenbestand, der von einer Suchmaschine aufgebaut wird.
- Kommando** s. Befehl
- Konkreter Informationsbedarf** Informationsbedarf, der sich auf ein Faktum richtet. Der Informationsbedarf ist mit der Anzeige des konkreten Faktums befriedigt.
- KWIC** s. Keywords in Context
- Laborstudie** Untersuchung (des Suchverhaltens), die in einem Labor mit Probanden stattfindet.

- Länderinterface** Auf das entsprechende Land angepasstes Interface einer Suchmaschine. Die Suche über verschiedene Länderinterfaces führt zu einer unterschiedlichen Reihung der Ergebnisse. S.a. Anfrageinterpretation.
- Linkkauf** Kauf von externen Links, um im Rahmen der Suchmaschinenoptimierung Linkpopularität zu gewinnen.
- Linksschiefe Verteilung** s. informetrische Verteilung
- Local Store** s. Datenbasis der Suchmaschine
- Logfile** Protokolldatei eines Webservers bzw. einer Suchmaschine, in der die technischen Interaktionen der Nutzer mit dem System festgehalten werden.
- Logfile-Analyse** Auswertung der Logfiles; in Suchmaschinen vor allem zur Gewinnung von Informationen über die Popularität von Suchbegriffen und das Interaktionsverhalten der Nutzer.
- Lokalität** Anpassung der Suchergebnisse (bzw. deren Reihung) an den momentanen Standort des Nutzers.
- Long Tail** s. Linksschiefe Verteilung
- Lookup searches** Suchanfragen, die dem Nachschlagen von Fakten oder der Befriedigung einfacher problemorientierter Informationsbedürfnisse dienen.
- Metadaten** Daten über andere Daten. Metadaten sind für Suchmaschinen vor allem bei der Indexierung nicht-textueller Inhalte und bei der Generierung von Trefferbeschreibungen relevant.
- Metasuchmaschine** Suchmaschine ohne eigenen Index, die die Ergebnisse mehrerer externer Suchmaschinen in einem eigenen Ranking zusammenfasst.
- Metatag** s. Metadaten
- Microblogging** Form des Bloggens, bei der die Länge der Nachrichten (künstlich) beschränkt ist.
- Mirror Site** dt. Spiegel: Kopie einer Website, die für einen schnelleren lokalen Zugriff angelegt wurde.
- N-Gramm** Ergebnis der Zerlegung eines Texts in Einheiten von N Zeichen.
- Need for freshness** Entscheidung, ob zu einer Suchanfrage aktuelle Dokumente von besonderer Bedeutung sind.
- Nicht eingreifende Methode** Untersuchungsmethode zur Beobachtung von Verhalten, ohne dass die Untersuchten von der aktuellen Beobachtung Kenntnis haben, zum Beispiel in Logfile-Untersuchungen.
- Nicht verbundene Seiten** Seiten, die zwar untereinander verlinkt sein können, jedoch von den einer Suchmaschine bereits bekannten Dokumenten aus nicht verlinkt werden.
- Nutzererleben** s. User Experience
- Oberflächenweb** Der Teil des Web, der von den allgemeinen Suchmaschinen erfasst werden kann.
- Online-Host** Anbieter von mehreren Datenbanken unter einer gemeinsamen Suchoberfläche.
- Online-Werbung** Werbung, die im Rahmen von Online-Angeboten geschaltet wird. Online-Werbung wird unterteilt in klassische Online-Werbung (z. B. Banner), Suchwortvermarktung und Affiliate-Werbung.

- Opaque Web** Undurchsichtiges Web; der Teil des Deep Web, der von den Suchmaschinen erfasst werden könnte, jedoch nicht erfasst wird.
- Operator** Spezieller Befehl zur Verbindung von Suchwörtern mit dem Ziel, die Menge der auszugebenden Treffer zu steuern. S.a. Boolesche Operatoren
- Organisches Ergebnis** Aus dem Web-Index automatisch generiertes Suchergebnis. Das Ranking der organischen Ergebnisse erfolgt für alle Dokumente zu gleichen Bedingungen, d. h. jedes Dokument, das in den Web-Index der Suchmaschine aufgenommen wurde, hat potentiell die gleiche Chance, als Ergebnis zu einer Suchanfrage angezeigt zu werden.
- Out-Link** Ein Link, der von einem bestimmten Dokument aus auf ein anderes Dokument zeigt.
- PageRank** Verfahren zur Bewertung von Dokumenten auf der Basis ihrer Verlinkung im Web. In diesem Verfahren wird sowohl die Zahl der eingehenden Links als auch ihr individueller Wert berücksichtigt.
- Paid Inclusion** Geschäftsmodell, bei dem Inhalteanbieter für die Aufnahme ihrer Inhalte in den Index einer Suchmaschine bezahlen.
- Parsing Module** Modul einer Suchmaschine zu Vorverarbeitung der vom Crawler gefundenen Dokumente.
- Partnerindex** Index, der von einer Suchmaschine mit eigenem Index Partnern zur Verfügung gestellt wird.
- Partnerindex-Modell** Modell, in dem eine Suchmaschine mit eigenem Index Suchergebnisse und Textanzeigen an einen Partner liefert, der zwar eine eigene Benutzerschnittstelle betreibt, jedoch die Ergebnisse des Partners anzeigt. Die aus den Anzeigenklicks erzielten Umsätze werden nach einem vorher festgelegten Schlüssel geteilt.
- Periodic Table of SEO Success Factors** Eine von der Website Search Engine Land erstellte Zusammenstellung der wichtigsten Rankingfaktoren der Suchmaschinen unter dem Blickwinkel der Suchmaschinenoptimierung.
- Power Law** s. infrometrische Verteilung.
- Precision** Anteil der relevanten Ergebnisse an der Gesamtzahl der von einer Suchmaschine angezeigten Ergebnisse.
- Precisiongraph** Darstellungsform für Precisionwerte, geordnet nach Trefferpositionen.
- Private Web** Seiten, die von den Autoren bewusst von der Indexierung durch Suchmaschinen ausgeschlossen wurden, beispielsweise durch Passwort-Abfragen.
- Problemorientierter Informationsbedarf** Informationsbedarf, dessen thematische Grenzen *nicht* exakt bestimmbar sind.
- Proprietary Web** Inhalte im Web, deren Nutzung nur nach Zustimmung zu bestimmten Nutzungsbedingungen möglich ist.
- Prosumer** Kombination aus consumer und producer, v. a. im Kontext von Social Media gebräuchlich.
- Qualitätsbestimmende Faktoren** Rankingfaktoren, mit denen die Qualität von Dokumenten gemessen werden soll, u. a. durch die Bewertung von Links und von Nutzerzugriffen.



**Quasisynonym** Wörter, die als Synonyme behandelt werden, obwohl sie keine sind. Suchmaschinen errechnen statistische Wortähnlichkeiten, die zur Behandlung von Wörtern als Quasisynonyme führen können.

**Quelltext** Text eines HTML-Dokuments inklusive der HTML-Befehle (Tags).

**Rankingfaktor** Kriterium, das zum Ranking von Dokumenten verwendet wird.

**Rankingsignal** s. Signal

**Recall** Anteil der von einem System gefundenen relevanten Dokumente an der Gesamtzahl der im Datenbestand vorhandenen relevanten Dokumente.

**Recht auf Vergessenwerden** Bezeichnung für ein durch die EU-Rechtsprechung gemachte Vorgabe, nach der bestimmte Informationen über eine Person auf Antrag dieser Person bei einer Suchmaschine (oder einem anderen Dienst) aus den Suchergebnissen entfernt werden müssen.

**Relevanz** Kriterium, ob ein Dokument von einem aktuell suchenden Nutzer in seinem Kontext als Suchergebnis gewünscht wird.

**Retrievaleffektivität** Die Fähigkeit einer Suchmaschine, auf eine Anfrage relevante Dokumente auszugeben

**Rich snippet** Erweiterte Trefferbeschreibung auf einer Suchergebnisseite.

**Robots Exclusion Standard** Von den allgemeinen Suchmaschinen akzeptierte Befehle, die die Steuerung der Suchmaschinen-Crawler und der Indexierung von Websites durch Suchmaschinen ermöglichen.

**Robots.txt** Datei, in der den Crawlern der Suchmaschinen Anweisungen zur Indexierung bzw. zum Ausschluss von Inhalten einer Website gegeben werden können.

**RSS** Format für die Übermittlung strukturierter Daten.

**Screenreader** Software, die die auf dem Bildschirm angezeigten Inhalte vorliest.

**Scroll area** below the fold

**Scrollen** Verschieben von Bildschirminhalten; durch die Bewegung der Inhalte lassen sich weitere Inhalte darstellen.

**Search Console** s. Google Search Console

**Search Engine Advertising (SEA)** s. Werbung

**Search Engine Marketing (SEM)** s. Suchmaschinenmarketing

**Search Engine Relationship Chart** s. Beziehungsgeflecht der Suchmaschinen

**Search Engine Result Page** s. Suchergebnisseite

**Seed Set** Sammlung von Websites bzw. Seiten, die als Startpunkte für das Crawling verwendet werden.

**Semantik** Lehre von der Bedeutung der Zeichen.

**SERP** s. Suchergebnisseite

**Server** s. Webserver

**Session** Abfolge von Suchanfragen und Dokumentsichtungen, die von einem bestimmten Nutzer innerhalb einer bestimmten Zeitspanne zu einem bestimmten Thema ausgeführt wurden.

**Share (in sozialen Netzwerken)** Teilen von Inhalten, um sie den mit dem Teilenden in einem sozialen Netzwerk verbundenen Personen sichtbar zu machen.

- Sichtbarer Bereich** Bereich der Suchergebnisseite, der ohne Scrollen sichtbar ist.
- Signal** Merkmal, das für das Ranking benutzt wird. Bei einem Signal handelt es sich um die kleinste Einheit, beispielsweise um die Feststellung, ob eine Website HTTPS verwendet. Dies wiederum ist ein Indikator für die Sicherheit dieser Website als Rankingfaktor, in den mehrere Signale eingehen können.
- Silo** Geschlossene Informationssammlung, die Externen (z. B. Suchmaschinen) nicht oder nicht vollständig zugänglich ist.
- Sitelinks** Links innerhalb der Trefferbeschreibungen auf Suchergebnisseiten, die direkt zu bestimmten Bereichen einer Website führen.
- Sitemap** Zusammenstellung von Links auf alle Seiten einer Website in einem Dokument; s.a. XML-Sitemap
- Social Media** Sammelbegriff für internet-basierte mediale Angebote, die auf sozialer Interaktion und den technischen Möglichkeiten des sog. Web 2.0 basieren.
- Spam** Von den Suchmaschinen 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.
- Spezielsuchmaschine** Suchmaschine, die sich thematisch oder anhand formaler Dokumentenmerkmale (Bsp. Dateityp) beschränkt.
- Spider** s. Crawler
- Spider Trap** Spider Traps sind Fallen, die dem Crawler gestellt werden, sei es absichtlich oder unabsichtlich.
- Spiegel** s. Mirror Site
- Sponsored Link** s. Textanzeige
- Standardsuche** s. Einfache Suche
- Stoppwort** Wort, das in der Indexierung ausgeschlossen wird.
- Streuverluste** Verluste bei der Schaltung von Werbung, die dadurch entstehen, dass die Zielgruppe nicht exakt erreicht werden kann.
- Suchanfrage** Eingabe eines Nutzers, die von einer Suchmaschine in einem Schritt verarbeitet wird. Eine Suchanfrage kann aus einem oder mehreren Wörtern, zusätzlich auch aus Operatoren und Befehlen bestehen.
- Suchbegriff** Einzelnes Wort, das Bestandteil einer Suchanfrage ist.
- Suchergebnisseite** Engl. *search engine results page (SERP)*: Zusammenstellung von Suchergebnissen auf Basis einer Suchanfrage. Zu einer Suchanfrage können mehrere Suchergebnisseiten ausgegeben werden, zwischen denen ein Nutzer blättern kann.
- Suchhistorie** s. Webprotokoll
- Suchmaschinenmarketing** Marketingmaßnahmen, die mittels Suchmaschinen durchgeführt werden.
- Suchmaschinenoptimierung** engl. *search engine optimization*; SEO: Maßnahmen, die bestimmten Dokumenten bzw. Websites zu einer höheren Sichtbarkeit in den Suchmaschinen verhelfen sollen.
- Suchverlauf** s. Webprotokoll

- Suchvokabular** Sammlung von Befehlen und Operatoren, die von einer Suchmaschine unterstützt werden.
- Suchvorschlag** Vorschlag zur Erweiterung oder Verbesserung einer Suchanfrage, der von einer Suchmaschine bereits während der Eingabe generiert wird.
- Suchwort** s. Suchbegriff
- Suchwortvermarktung** Bestandteil der Online-Werbung, bei dem Textanzeigen zu einzelnen Suchbegriffen an Werbetreibende verkauft werden.
- Surface Web** s. Oberflächenweb
- Synonym** Wörter, die einen gleichen oder sehr ähnlichen Bedeutungsumfang haben.
- Tab blindness** Bezeichnung für die Tatsache, dass Nutzer die von einer Suchmaschine angebotenen Verweise auf weitere Kollektionen, in denen sich spezialisierte Suchen durchführen ließen, übersehen.
- Tag** Auszeichnung von Daten mit zusätzlichen Informationen.
- Termfrequenz** Häufigkeit des Vorkommens von Termen (Wörtern oder Wortformen) in Dokumenten.
- Textanzeige** Anzeige auf der Ergebnisseite einer Suchmaschine, die als Antwort auf eine Suchanfrage ausgegeben wird und in ihrer Darstellung einem organischen Suchergebnis ähnelt.
- Textstatistik** Verfahren zur Auswertung von Texten, um auf der Basis von Worthäufigkeiten, Wortpositionen und weiteren Merkmalen ein Ranking zu erstellen.
- Toolbar** Browsererweiterung, die stets im Browserfenster sichtbar ist und dem Browser zusätzliche Funktionen, z. B. Suchfunktionen, hinzufügt.
- Topic detection and tracking** Zusammenfassen thematisch ähnlicher Dokumente zu einem Cluster und beständiges Hinzufügen neuer Dokumente zu passenden Clustern.
- Traffic** Besucher, die auf ein Dokument bzw. eine Website gelangen.
- TREC** Text Retrieval Conference: Evaluierungsinitiative für den Vergleich von Information-Retrieval-Systemen. Im Rahmen von TREC werden u. a. Testkollektionen entwickelt und bereitgestellt.
- Trefferbeschreibung** engl. *snippet*: Kurze Beschreibung eines Dokuments auf der Suchergebnisseite.
- Trefferliste** Liste der zu einer Suchanfrage aufgefundenen Dokumente. Eine Suchergebnisseite kann mehrere Trefferlisten sowie andere Darstellungsformen für Treffer enthalten.
- Truly Invisible Web** Seiten bzw. Websites, die für die Suchmaschinen aufgrund technischer Gegebenheiten nicht indexierbar sind.
- Trunkierung** Abschneiden eines Suchbegriffs durch Setzen eines Sonderzeichens (meist \*); sorgt dafür, dass mehr Dokumente gefunden werden. Bei der gängigen Rechtstrunkierung werden alle Wörter gefunden, die einen eingegebenen Wortstamm enthalten (so findet etwa *suchma\** neben Suchmaschine auch Suchmaschinen, Suchmaschinenmarketing, Suchmaschinenoptimierung, usw.).
- Umgebungstext** Text, der bei Multimedia-Inhalten (Bilder, Videos) auf HTML-Seiten in der Nähe des jeweiligen Objekts platziert ist. Umgebungstexte lassen sich für die Indexierung der Objekte nutzen.

**Universal Search** Anreicherung der Suchergebnisse aus dem Web-Index um Ergebnisse aus speziellen Kollektionen und deren gemeinsame Darstellung auf einer Suchergebnisseite.

**Unsichtbarer Bereich** Teil der Suchergebnisseite, der erst durch Scrollen sichtbar wird.

**Unsichtbares Web** s. Deep Web

**URL** Uniform Resource Locator: Eindeutige „Adresse“ eines Dokuments im Web.

**User Agent** Programm, mit dem Dienste im Internet genutzt werden, beispielsweise ein Web-Browser. Der User Agent überträgt in der Regel seinen Namen, wenn er eine Anfrage an einen Dienst stellt. Die Crawler der Suchmaschinen sind auch User Agents und identifizieren sich entsprechend. Sie können damit von Websitebetreibern auch erkannt werden.

**User Experience** Alle Aspekte der Erfahrung eines Nutzers mit einer Suchmaschine.

**Vertikale Suchmaschine** s. Spezialsuchmaschine.

**Verweildauer** Engl. *dwell time*: Zeit, die Nutzer mit dem Ansehen eines Dokuments verbringen. Kann bei einer bestimmten Länge als Qualitätsindikator für das Dokument angesehen werden kann, bei sehr kurzer Dauer (und einem Zurückkehren auf die Trefferliste der Suchmaschine aber als Indikator gegen die Qualität des Dokuments verwendet wird.

**Verzerrungen** s. Bias

**Visible area** s. Sichtbarer Bereich.

**Web** s. World Wide Web

**Web 2.0** s. Social Media

**Web-Verzeichnis** Von Menschen erstellte Zusammenstellung von Websites. Die Websites werden in ein Klassifikationssystem eingeordnet und mit kurzen Texten beschrieben.

**Webmaster Tools** s. Google Search Console

**Webprotokoll** Die von einer Suchmaschine gesammelten Suchanfragedaten eines einzelnen, identifizierbaren Nutzers.

**Webseite** auch Seite; engl. *web page*: Ein einzelnes Dokument, das sich im Web über eine eindeutige URL erreichen lässt.

**Webserver** Computer, auf dem eine Website abgelegt ist.

**Website** engl. *web site*: Ein Angebot im Web, welches mehrere Dokumente („Seiten“) enthält.

**Werbung** Im Kontext von Suchmaschinen vor allem zu einer Suchanfrage passende Textanzeigen, die auf den Suchergebnissen gezeigt werden.

**White Hat** In der Suchmaschinenoptimierung Maßnahmen, die mit den von den Suchmaschinenbetreibern aufgestellten Regeln konform sind.

**White List** Zusammenstellung von durch Menschen ausgewählter Websites, die beispielsweise für das fokussierte Crawling oder als *seed set* für das allgemeine Crawling eingesetzt werden können.

**WHOIS-Service** Verzeichnis von Inhaberdaten zu Websites.

**World Wide Web** Dienst des Internet, der aus miteinander verbundenen Hypertext-Dokumenten besteht. Jedes Dokument besitzt eine einzigartige Adresse (Uniform Resource

Locator; URL), mit der es von anderen Dokumenten aus adressiert werden kann. Die Verbindungen zwischen Dokumenten heißen Links.

**WWW** s. World Wide Web

**Wortstamm** Bestandteil eines Worts, der als Ausgangsbasis für Wortbildungen dienen kann.

**XML** Extensible Markup Language; Textauszeichnungssprache, die die Auszeichnung hierarchisch strukturierter Daten erlaubt.

**XML-Sitemap** Datei, die es Websitebetreibern erlaubt, den Crawlern der Suchmaschinen detaillierte Empfehlungen für das Crawling vor allem großer Websites zu geben.

---

# Stichwortverzeichnis

## A

Accuracy (Sorgfalt), 234  
Adresszeile, 13  
AdSense, 194  
AdWords, 135  
AGOF (Arbeitsgemeinschaft Online  
Forschung), 3  
Aggregatoren, 257  
Ähnliche Seiten (Option), 152  
Aktualisierungsdatum, 118  
Aktualisierungsfrequenz, 118  
Aktualität, 269  
Algorithmen  
  Objektivität, 210  
Algorithmische Suchmaschine, 17  
allgemeine Suchmaschine, 17  
AltaVista, 18, 164  
alternative Suchmaschine, 201  
  Gründe für die Nutzung, 201, 203  
Amazon, 3, 110  
Ambient Findability, 304  
Analysedienst, 115  
Anfrage-Dokumenten-Paradigma, 306–307  
Anfrageinterpretation, 59  
  explizite, 62  
  implizite, 62  
Anfragetyp, 13, 205  
  informationsorientierter, 71, 206, 216, 241,  
  278  
  navigationsorientierter, 71, 205, 215, 241,  
  278  
  transaktionsorientierter, 72, 206, 216  
  Unterscheidung der Anfragetypen, 71

  Verteilung der Anfragetypen, 73  
Angemessenheit, 234  
Ankertext, 56, 59, 95  
Anzeigen Siehe Werbung  
AOL, 164  
API (Application Programming Interface), 160  
Application Programming Interface, 160  
Arbeitsgemeinschaft Online Forschung  
  (AGOF), 3  
Archivsuchmaschine, 266  
ARD/ZDF-Onlinestudie, 3  
Ask.com, 164  
auslösendes Ereignis, 85  
authority (Autorität), 176, 180  
Autocomplete (Suchvorschläge), 78  
Autorität, 176, 180

## B

Baidu, 162, 164  
Barrierefreiheit, 174  
Befehle  
  filetype, 86  
  Verwendung durch Nutzer, 86  
Befragung, 69  
Benutzerführung, 207  
Benutzerschnittstelle, 30  
bestätigende Information, 234  
Bias Siehe Suchmaschinen-Bias  
Bildersuchmaschine, 263  
Bildschirmauflösung, 132  
Bing, 159, 163, 207, 257  
  erweiterte Suche, 225  
iPad App, 304

- Marktanteil, 161
- Retrieval-effektivität, 239
- Bing Maps, 291
- black hat, 182
- Blog, 34
- Boolesche Operatoren, 86, 219
  - Beispiele für die Verwendung, 222
  - Klammersetzung, 221
  - Verbindung von Suchanfragen, 223
  - Verwendung bei Google, 222
- bow-tie model (Fliegen-Modell), 34
- Breadcrumb-Navigation, 150
- Browsing, 75
  
- C**
- Cache, 152
- Cache-Kopien, 40
- CARS-Checkliste, 234
- Clewwa, 266
- click fraud (Klickbetrug), 195
- Cloaking, 179, 182
- Cluster
  - Google Scholar, 274
  - Nachrichtensuche, 270
- Content Acquisition, 36
- Country Bias, 39
- Crawler, 37
  - für spezielle Kollektionen, 48
- Crawling, 30, 36–37, 178
  - Ausschluss von Inhalten, 41, 44
  - fokussiertes, 48, 266
  - Prioritäten, 38
  - Probleme, 39
  - Steuern, 41
- Crawling-Frequenz, 48
- credibility (Glaubwürdigkeit), 234
  
- D**
- Dark Web, 253
- Dateiformat, 254
- Datenbank, 21
  - im Deep Web, 254
- Datenbasis der Suchmaschinen, 46
- Dauerbrenner, 85
- Deep Web, 38, 251, 275, 285
  - Bedeutung, 261
  - Bedeutung für die Recherche, 260
  - Definition, 252
  - Größe, 259
- Multimedia-Inhalte, 255
- Problem der allgemeinen Suchmaschinen, 264
  - Typen von Inhalten, 253
- dialogorientierte Suche, 77
- direkte Antwort, 146
- disconnected pages (nicht verbundene Seiten), 254
- DMOZ-Verzeichnis (Open Directory Project), 24, 257
- Dokument, 21, 34
  - Repräsentation, 32, 50, 56, 58, 146
  - Verstehen der Dokumente, 306
- Dokumenttyp, 266
- Domainhoster, 235
- Dublette, 39, 178, 191, 206, 244, 259, 275
- duplicate content *Siehe Dublette*
- dwell time (Verweildauer), 60, 111, 246
- dynamisch generierte Inhalte, 254
  
- E**
- Ebay, 3
- Echtzeit-Inhalte, 255
- Einschränkung der Treffermenge, 209
- Endgerät
  - Einfluss auf die Ergebnisdarstellung, 132
- Erfassung von Nutzerdaten
  - Analysedienste für Websitebetreiber, 115
  - eigene Browser, 115
  - Logfiles, 113
  - Personalisierungstools, 115
  - Toolbars, 114
- Ergebnispräsentation, 131, 136
- Erstellungsdatum, 118
- erweiterte Suche, 224
- Erweiterung der Treffermenge, 209
- eSearch plus, 257
- Evaluierung der Suchergebnisse, 239
- Event, 84
- Evergreen (Dauerbrenner), 85
- Exalead, 164
- explorative Suche, 75
- exploratory search (explorative Suche), 75
- Eyetracking, 138, 154
  
- F**
- F pattern (F-Muster), 139
- F-Muster, 139
- Facebook, 105, 110, 282, 284

Faktenextraktion, 146  
 Fakteninformationen, 135  
 Faktensuchmaschine, 275  
 Feeds, 36  
 filetype, 86  
 Filter Bubble (Filterblase), 124  
 Filterblase, 124  
 Firefox, 162  
 Firmenwissen.de, 258  
 Fixation, 138  
 Flash, 178  
 Fliegen-Modell, 34  
 focused crawling (fokussiertes Crawling), 48, 266  
 fokussiertes Crawling, 48, 266  
 Folksonomy, 25  
 Frage-Antwort-Dienst, 25  
 FragFinn, 21

## G

Genios, 258  
 Geschäftsmodell der Suchmaschinen, 159  
 gesellschaftliche Bedeutung der Suchmaschinen, 3, 289  
 Glaubwürdigkeit, 234  
 Gmail, 115  
 golden triangle (Goldenes Dreieck), 138  
 Goldenes Dreieck, 138  
 Google, 6, 62, 104, 118, 163, 173, 208, 289  
   als Anbieter von Inhalten, 212  
   als Informationsvermittler, 212  
   als Unternehmen, 3, 159  
   Alternativen, 201, 209  
   Bevorzugung eigener Angebote, 210  
   Boolesche Operatoren, 222  
   Erfassung von Nutzerdaten, 115  
   erweiterte Suche, 224  
   Kollektionen, 47  
   Kombination von Booleschen Operatoren, 228  
   Kooperation mit der Mozilla-Foundation, 162  
   Marktanteil, 161  
   nicht mögliche Suchanfragen, 228  
   Retrievaleffektivität, 239  
   Spezialsuchmaschinen, 210  
   Übersicht der Crawler, 48  
   Updates, 185

  Verzerrungen, 210  
 Google Analytics, 115, 185  
 Google Bildersuche, 225  
 Google Books, 211, 274  
 Google Chrome, 115  
 Google Custom Search, 267  
 Google Maps, 291  
 Google News, 21, 49, 266, 268, 304  
   Clusterbildung, 270  
   Quellen, 269  
 Google Places, 291  
 Google Scholar, 258, 272  
   Dokumenttypen, 273  
   Fehler bei der Indexierung, 273  
   Inhalte, 272  
 Google Search Appliance, 160  
 Google Shopping, 160, 211  
   Paid Inclusion, 300  
 Google Street View, 211  
 Google Translate, 160  
 Google Trends, 84  
 Google Übersetzer, 151  
 Google Videos, 31  
 Google Webmaster Tools, 185  
 Google Zeitgeist, 84  
 Google+, 115, 211  
 Google-Account, 123  
 Google-Gesellschaft, 3  
 GPS, 120

## H

Hard Coding Bias, 212  
 Heatmap, 138, 154  
 hidden text *Siehe Suchmaschinenoptimierung*  
 HITS (Hypertext Induced Topic Search), 105  
 HTML-Tags, 179  
 Hybridsuchmaschine, 21, 257, 266, 272  
 Hypertext Induced Topic Search, 105

## I

IDF (Inverse Dokumenthäufigkeit), 100  
 Impuls, 85  
 In-Link, 104  
 Index, 49  
   Ausschluss von Inhalten aus dem Index, 44  
   invertierter Siehe invertierter Index, 49  
 Indexer, 49  
 Indexierung, 49



- Bilder, 55
- Strukturinformationen, 54
- Umgebungstexte, 55
- Videos, 56
- Information Broker, 261
- Information-Retrieval-Evaluierung, 239
- Informationsbedarf, 68, 78, 137, 242–243
  - konkreter vs. informationsorientierter, 70
- Informationskompetenz, 80, 310
- informatrische Verteilung, 82, 99, 259
- Inhalte der Suchmaschinen, 5
- interaktives Information Retrieval, 246
- inverse Dokumenthäufigkeit, 100
- inverted document frequency (inverse Dokumenthäufigkeit), 100
- invertierter Index, 49, 96
  - einfacher, 51
  - mit Positionsangaben, 52
  - mit Worthäufigkeiten, 52
- invisible Web *Siehe Deep Web*
- IP-Adresse, 120
  
- J**
- Juroren, 243
  
- K**
- keyword density (Keyworddichte), 99
- keyword stuffing, 126, 179, 182
- Keyworddichte, 99
- Keywords in Context, 150
- Klickbetrug, 195
- Knowledge Graph, 135, 137, 144, 235, 277, 306–307
- Kollektion, 46, 135, 141–142, 209, 211
  - Aufnahmekriterien, 144
- Kontext der Suche, 16
- Kontextualisierung, 123
  
- L**
- Laborstudie, 69
- Ladegeschwindigkeit, 125, 178
- Länderinterface, 122
- Lexis Nexis, 257
- Link
  - Qualität, 104
  - unidirektional vs. bidirektional, 105
- Linked Open Data, 306
- LinkedIn, 284
- Linkkauf, 104, 182
- linktopologisches Ranking-Verfahren, 102
- Logfile, 245
- Logfileuntersuchung, 69
- lokale Suche, 143
- Long Tail, 83
- lookup searches, 74
- Lycos, 18
  
- M**
- Manipulation der Suchergebnisse, 5, 175
- Meinungsvielfalt, 204
- Meta Description, 150
- Metadaten, 42, 57
  - aus dem Dokument extrahierte, 58
  - aus dem Web, 58
  - dem Dokument beigegebene, 58
  - über die Website, in der ein Dokument abgelegt ist, 58
- Metager, 23
- Metasuchmaschine, 22
  - Kritik, 22
  - Ranking, 23
- Metatag, 150
  - Description, 179
  - title, 100
  - Überschriften, 100
- Microblogging, 283
- Mirror Site, 19
- mobile Suche, 16
- Mozilla, 162
- Multimedia-Inhalte im Deep Web, 255
  
- N**
- Nachrichtensuche, 142
- Nachrichtensuchmaschine, 46, 268
  - Aktualität, 269
  - Beschränkung auf kostenlose Inhalte, 269
- Navigationselemente auf der Suchergebnisseite, 147
- Need for freshness, 118
- neutrale Suchmaschine, 292
- nicht reaktives Verfahren, 69
- nicht verbundene Seiten, 254
- Nutzermodell
  - Spezialsuchmaschinen, 265
  - Suchmaschinen, 265
- Nutzerverhalten, 67

Methoden zur Erfassung, 69  
 Nutzung von Suchmaschinen, 3  
 nutzungsstatistische Verfahren, 109  
   Auswertung von Klicks auf Suchergebnis-  
   seiten, 110  
   Bewertung, 116  
   Datenerhebung, 112  
   Ebenen der Datenerfassung, 113  
   Ebenen der Datenerhebung, 109  
   explizite vs. implizite Bewertungen, 110  
   Klicks als Stimmen für ein  
   Dokument, 111

**O**

Oberflächenweb, 38, 253, 268  
 Online-Werbung, 161  
 Opaque Web, 260  
 Open Directory Project, 24, 257  
 Operatoren, 86  
   Verwendung, 86  
 organische Ergebnisse, 133, 140  
 Out-Link, 104

**P**

PageRank, 95, 104  
   als Mythos, 108  
   Berechnung, 106  
   Dämpfungsfaktor, 106  
   Einsatz in der Praxis, 105  
   Gewichtung von Links, 106  
   historische Bedeutung, 107  
   Iteratives Verfahren, 107  
 paid inclusion, 160, 300  
 Paperball, 268  
 Paperboy, 268  
 parsing module (Syntaxanalyse), 49  
 Partnerindex-Modell, 164  
 Periodic Table of SEO Success Factors, 175  
 Personalisierung, 116, 123  
   nutzbare Daten, 124  
   personenbezogene Faktoren in der Suchma-  
   schinenoptimierung, 181  
   Steuerung durch den Nutzer, 123  
 Personalisierungstool, 115  
 Phrase, 53  
 Phrasensuche, 87, 224  
 Popularität, 102  
   nach Nutzergruppen, 102

Power Law, 83  
 Precision, 217, 239, 242  
 Precisiongraph, 239  
 private Web, 260  
 proprietary Web, 260  
 prosumer, 109

**Q**

Qualität  
   Abbildung durch Popularität, 103  
 Qualitätsmessung, 248  
 Quellenauswahl, 217  
 Quellenprüfung, 235  
 Query understanding *Siehe*  
   *Anfrageinterpretation*

**R**

random surfer (Zufallssufer), 105  
 Ranking, 30, 93  
   Adaptierbarkeit der Dokumente für mobile  
   Endgeräte, 125  
   Aktualität, 95, 117, 177  
   Annahmen, 62  
   Anzahl der Faktoren, 94  
   Einfluss der Position auf das Selektionsver-  
   halten, 154  
   Ermittlung potentiell relevanter Dokumente,  
   97  
   formale Qualitätsbewertung vs. Vertrauens-  
   würdigkeit, 238  
   Gleichbehandlung der Dokumente, 138  
   Grundannahmen, 93  
   linktopologische Verfahren, 102, 104, 108  
   Lokalität, 119, 122, 181  
   Nachrichtensuchmaschinen, 269  
   nutzungsstatistische Verfahren, 103  
   Personalisierung, 95  
   Popularität, 95, 102  
   Qualitätsmessung, 95  
   technische Faktoren, 96, 125  
   textspezifische Faktoren, 95  
   Textstatistik, 96  
   Textstatistik als Basis, 97  
   Umgang mit Spam, 127  
   Worthäufigkeit, 98  
 Rankingupdates, 182  
 reasonableness (Angemessenheit), 234  
 reasoning, 307

- Recall, 217, 240
- Recherche
  - Auswahl der Suchbegriffe, 219
  - Befehle, 226
  - Befehle vs. erweiterte Suchformulare, 226
  - Deep Web, 260
  - Effizienz, 230, 261
  - Effizienz vs. Effektivität, 218
  - Einschränkungsmöglichkeiten in der erweiterten Suche, 224
  - Kombination von Booleschen Operatoren bei Google, 228
  - komplexe Suche, 229
  - Kosten, 261
  - Suchanfragen, die man bei Google nicht stellen kann, 228
  - Suche nach in einem Dokument enthaltenen Mediendateien, 229
  - Zugänglichkeit im Web vs. über das Web, 256
- Recht auf Vergessenwerden, 40
- Reihungseffekt, 111
- Relevanz, 30
- Relevanzurteile
  - explizite vs. implizite, 246
- Retrievaleffektivität, 239, 242
  - sessionbasierte Evaluierung, 246
- Retrievaltest, 242
  - Anzahl der Suchanfragen, 243
  - Nutzergruppen, 243
  - Standardaufbau, 242
  - Testaufbau, 242
  - Verwendung von Klickdaten, 245
- Rich snippet, 152
- robots.txt, 42
  
- S**
- schema.org, 152, 306
- SEA (search engine advertising), 172
- search engine advertising, 172
- search engine marketing (Suchmaschinenmarketing), 172
- search engine optimization *Siehe Suchmaschinenoptimierung*
- search engine results page *Siehe Suchergebnisseite*
- search gap, 297
- Searcher, 59
- seed set, 37
- Selektionsverhalten, 139, 154
- SEM (Suchmaschinenmarketing), 172
- semantic Web, 61
- Semantic Web, 306
- Semantik, 61
- semantische Suche, 61
- SEO *Siehe Suchmaschinenoptimierung*
- SERP *Siehe Suchergebnisseite*
- Session, 73
  - Länge, 74
- Siri, 163
- snippet *Siehe Trefferbeschreibung*
- Social Media, 109, 180, 281
  - als Rechercheinstrumente, 284
  - als Recherchequellen des Deep Web, 284
  - bidirektionale Verbindung, 284
  - im Kontext der Suchmaschinenoptimierung, 182
  - Influencer, 182
  - Sichtbarkeit in Suchmaschinen, 281
  - Typen, 282
- Social-Bookmarking-Dienst, 25
- Sorgfalt, 234
- Soziale Medien *Siehe Social Media*
- soziale Netzwerke, 124, 283
- soziale Suche, 286
- Spam, 40, 102, 126, 180–181, 257
  - Erkennung im Crawling, 126
  - Suchmaschinenoptimierung, 182
  - Umgang mit Spam im Ranking, 127
- Spezialsuchmaschine, 19, 47, 142, 263, 291
  - als Quelle, 218
  - Beschränkung auf einen Dokumenttyp, 266
  - Typen, 266
- Spider Trap, 40
- Spider *Siehe Crawler*
- sponsored link *Siehe Suchmaschinenwerbung*
- Spotify, 308
- Standardsuchmaschine
  - im Betriebssystem, 162
  - im Browser, 162
- Standort
  - eines Dokuments, 120
  - eines Nutzers, 120
- Steuerung der Ergebnismenge, 215
- Stoppwort, 100
- Streuverluste, 4

- Strukturinformationen, 100
- Suchanfragen, 75
  - Anzahl der Suchanfragen in Retrievaltests, 243
  - durchschnittliche Länge, 81
  - Einwort-Anfragen, 81
  - Formulierung, 78, 215
  - Fragesätze, 82
  - Häufigkeiten, 82
  - Länge, 80
  - Mehrwortbegriffe, 81
  - Möglichkeiten, 75
  - navigationsorientierte, 83
  - Suchvolumen im zeitlichen Verlauf, 84
  - Themen, 87
  - Trends, 84
  - Veränderung, 147
  - Verstehen der Suchanfragen, 306
  - Verteilung, 81, 83
  - Volumen, 161
- Suche als Basistechnologie, 304
- Suchergebnisse
  - Beeinflussung der Trefferpositionen, 292
  - Ebenen der Qualitätsprüfung, 234
  - Evaluiierung, 239
  - Glaubwürdigkeit, 289
  - ideale Menge, 292
  - interessengeleitete Präsentation, 296
  - Kriterien für die Bewertung der Qualität, 233
  - Manipulation, 5
  - menschliche vs. maschinelle Qualitätsüberprüfung, 235
  - Neutralität, 291
  - organische, 296
  - Qualität, 233
  - Trennung von Werbung und organischen Ergebnissen, 308
  - Überschneidung, 202
  - Unzufriedenheit, 208
  - Vertrauenswürdigkeit, 238
  - Verzerrung durch Suchmaschinenoptimierung, 173
  - wissenschaftliche Evaluierung, 239
- Suchergebnisseite, 131
  - Anzahl der Treffer, 138
  - Aufbau, 133
  - Elemente, 140
  - Optionen, 148
  - Optionen zu den einzelnen Suchergebnissen, 152
  - Selektionsverhalten, 154
  - sichtbarer vs. unsichtbarer Bereich, 138
- Suchhistorie, 60
- Suchmaschine
  - als Anbieter von Informationsinfrastrukturen, 309
  - als Black Box, 94
  - als Datensammler, 309
  - als Informationsvermittler mit eigenen Interessen, 308
  - als Inhalteanbieter, 291
  - als Public Service, 309
  - Beschränkungen, 264
  - Datenbasis, 46
  - Definition, 29
  - grundlegender Aufbau, 31
  - Index vs. Suchmaschine, 164
  - Inhalte, 36
  - Interessen der Suchmaschinenbetreiber, 290
  - neutrale, 292
  - Nutzung, 3
  - Probleme der Erschließung, 265
  - Rolle als Vermittler von Informationen, 289
  - Rolle für den Wissenserwerb, 308
  - Sichten auf die Inhalte des Web, 204
- Suchmaschinen-Bias, 292
  - Auswirkungen, 293
  - Probleme, 293
- Suchmaschinenbetreiber
  - Position, 184
  - Verantwortung, 291
- Suchmaschinenmarketing, 172
- Suchmaschinenmarkt, 159
  - Gewinnung von Marktanteilen, 162
  - Markteintrittsbarrieren, 163
  - Marktverhältnisse als Ausdruck von Qualität, 201
  - Marktanteile, 161
- Suchmaschinenoptimierung, 141, 171
  - Alter der Domain, 180
  - Autorität, 176
  - Bedeutung, 172, 290
  - für informierende Inhalte, 174
  - für vertikale Suchmaschinen / Universal Search, 183

- grundlegende Verfahren, 174
  - Hilfestellungen der Suchmaschinenbetreiber, 185
  - Inhalte, 177
  - Linkaufbau, 180
  - Linkkauf, 181
  - Off-The-Page-Optimierung, 180
  - On-The-Page-Optimierung, 176
  - Optimierung der HTML-Struktur, 179
  - personenbezogene Faktoren, 181
  - Position der Suchmaschinenbetreiber, 184
  - pragmatischer Ansatz, 175
  - Public Relations, 174
  - Social Media, 182
  - Spam, 182
  - Unterscheidung zwischen On-The-Page-Faktoren und Off-The-Page-Faktoren, 175
  - versteckter Text, 179
  - Vertrauen, 180
  - white hat vs. black hat, 182
  - Suchmaschinentechnologie, 29
    - Verkauf als Erlösmodell, 160
  - Suchmaschinenwerbung, 4, 189
    - als Suchergebnis, 189
    - Eigenschaften, 193
    - Gestaltung der Anzeigen, 194
    - innerhalb der Universal-Search-Ergebnisse, 197
    - Kennzeichnung auf den Suchergebnisseiten, 196
    - Kennzeichnung der Anzeigen, 196
    - Qualitätsfaktoren, 195
    - Ranking, 193
    - Ranking der Anzeigen, 195
    - Unterscheidbarkeit von Anzeigen und organischen Ergebnissen, 191, 196
  - Suchneutralität, 292, 299
  - Suchprozess, 67
    - Entscheidungen, 68
  - Suchvorschläge, 78
  - surface web (Oberflächenweb), 38, 253, 268
  - Synonym, 219
    - Suchanfrage, 222
  - Syntaxanalyse, 49
- T**
- T-Online, 164
  - Tab blindness, 142, 278
  - Taxonomie der Online-Informationen, 256
  - Tendrils, 35
  - Termfrequenz, 100
  - Textanzeigen, 135, 141
  - Textanzeigen Siehe Suchmaschinenwerbung
  - Textstatistik, 269
  - TF\*IDF (term frequency \* inverted document frequency), 100
  - Toolbars, 114
  - Topic Detection and Tracking, 270
  - Traffic, 162, 172, 281
    - als Rankingfaktor, 118
  - TREC, 242
  - Trefferbeschreibung, 149
    - Aufbau, 149
    - Basisinformationen, 150
    - im Kontext von Retrievaltests, 244
    - Quelle des Beschreibungstexts, 150
    - Varianten, 151
  - Trefferliste, 131
  - trigger word, 120
  - Trivia, 216
  - Truly Invisible Web, 260
  - tubes, 35
  - Twitter, 284
- U**
- Überschneidungen zwischen Suchergebnissen, 202
  - Umgebungstext, 55
  - Universal Search, 46, 263
    - Einbindung von Spezialsuchmaschinen, 277
  - Universal-Search-Ergebnisse, 135, 137, 139, 142, 207, 218, 296
  - Universalsuchmaschine Siehe Web-Suchmaschine
  - Unsichtbares Web Siehe Deep Web
  - Unzufriedenheit mit den Suchergebnissen, 208
- V**
- Veränderungsfrequenz, 58
  - Vertikale Suchmaschine Siehe Spezialsuchmaschine
  - verwandte Begriffe, 219
  - Verweildauer, 60, 111, 246
  - Verzerrung der Suchergebnisse durch Suchmaschinenoptimierung, 173

**W**

- Wayback Machine, 267
- Web Crawling Siehe Crawling
- Web-Index, 46
- Web-Katalog, 23
- Web-Suchmaschine, 17
  - als Suchmaschinentyp, 17
- Web-Verzeichnis, 23
- Web.de, 164
- Webcrawler, 18
- Webprotokoll, 114–115
- Website-Architektur, 178
- Websites (Siehe auch Dokument), 20
  - beliebteste, 3
- Werbung, 135–136, 141, 159, 296
  - als relevantes Suchergebnis, 142
  - search engine advertising, 172
  - Teilen der Erlöse im Partnerindex-Modell, 164
  - Unterscheidbarkeit von den organischen Suchergebnissen, 296
- white hat, 182
- white list, 266, 269
- WHOIS, 235
- wiederkehrendes Ereignis, 84
- Wikipedia, 103, 146, 238, 307
- Wissenschaftssuchmaschine, 272
- Wolfram Alpha, 62, 275
- World Wide Web
  - Erfassung durch Suchmaschinen, 32
  - In-Bereich, 35
  - Kern, 35
  - Out-Bereich, 35
  - verteilte Inhalte, 29

**X**

- Xing, 284
- XML-Sitemap, 42, 44

**Y**

- Yahoo, 160, 164, 207, 225
- Yahoo Clever, 26
- Yandex, 162, 164
- YouTube, 31, 144, 211, 291

**Z**

- Zensur, 44
- Zufallssurfer, 105