

Suchen, um zu finden

Steffen-Peter Ballstaedt



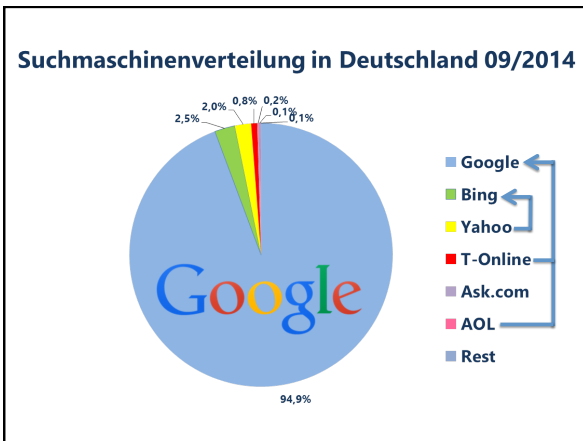
24.6.2015



Statistik BRD

(Quelle: Comscore, 2013)


- etwa 6 Milliarden Suchanfragen pro Monat
- etwa 2.500 Anfragen pro Sekunde
- 96% der Suchanfragen über Google

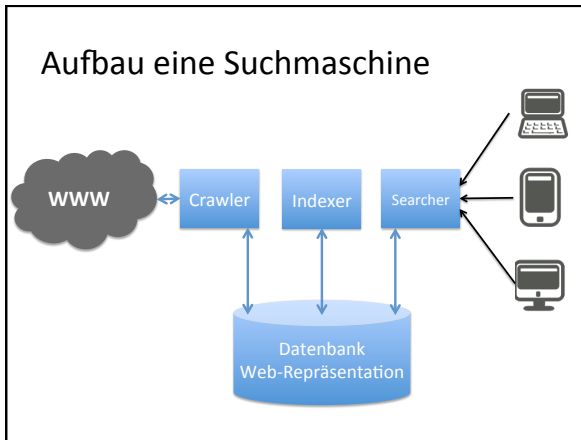


Antworten auf folgende Fragen:

- Wie findet eine Suchmaschine Inhalte bzw. Dokumente im WWW?
- Wie kommt die Trefferliste zustande (Ranking)?
- Was macht geschicktes und professionelles Suchen aus?
- Welche Alternativen zu Google gibt es?
- Wie kann man die Seriosität der Treffer beurteilen?

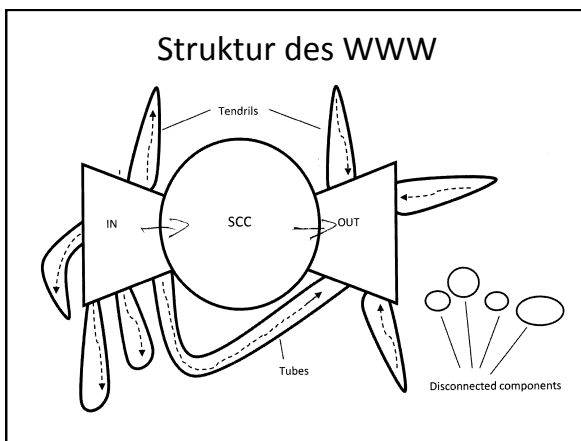
Wie findet eine Suchmaschine Inhalte bzw. Dokumente im WWW?





Crawling

- Ausgehend von bekannten Websites werden die Links verfolgt.
- Die Dokumente werden erfasst und gespeichert.
- Es gibt Crawler für bestimmte Inhalte (Kollektionen).
- Crawling findet kontinuierlich statt, nach Aktualität und Popularität



Deep Web

- Webseln ohne Links zu ihnen (Dark Web)
- Durch Passwort geschützte Bereiche
 - Beispiel: Intranets
- Suchformulare für Datenbanken
 - Beispiel: Telefonbuch
- Social-Media-Dienste
 - Beispiel: Facebook, Twitter
- nicht erfassbare Datei-Formate
 - Beispiel: Flash-Dateien
- Echtzeitinhalte, die sich ständig verändern
 - Beispiele: Börsenkurse, Wetterdaten
- Inhalte, die für Crawler bewusst gesperrt sind

Indexierung

- Erstellt eine Repräsentation des Dokuments aus sprachstatistischen Daten: Worthäufigkeit, Wortposition, Wort in Titel oder Überschrift.
- Der Index als eine riesige Datenbank ist das Herz einer Suchmaschine.
- Bei einer Suchanfrage wird nicht auf die Dokumente im Web, sondern auf diese Repräsentation zugegriffen.

Searcher

- Interpretation der Suchanfrage (der eingegebenen Stichwörter):
 - Suchhistorie des Nutzers
 - Suchanfrage in der jeweiligen Session
 - angeklickte Dokumente
 - Verweildauer auf einem Treffer
 - Suchverhalten vergleichbarer Nutzer

Wie kommt die Trefferliste zustande (Ranking)?

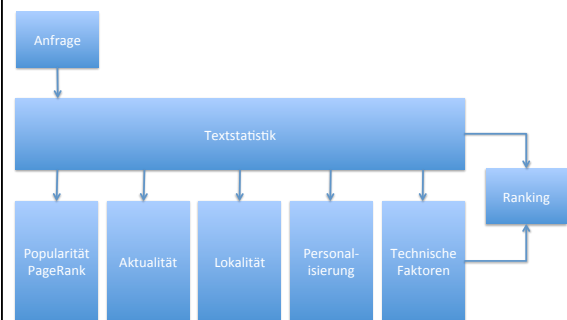


Ranking

- Die Treffer sind nach absteigender Relevanz aufgelistet.
- Das Ranking wird als Kombination und Gewichtung verschiedener Faktoren berechnet (200 bei Google)
- Die Berechnung des Ranking ist bei jeder Suchmaschine anders und ein Geheimnis (Algorithmus)!
- Vergleich von Rankings



Erzeugung der Trefferliste



Ranking-Kriterien

- In das Ranking gehen keine qualitativen, sondern nur quantitative Größen ein:
 - linkstatistische Variablen
 - nutzerstatistische Variablen
- Suchmaschinenoptimierer versuchen, Ihre Institution möglichst oben in der Trefferliste zu platzieren.

Diese Daten speichert Google



- die IP-Adresse des Nutzers
- die Suchanfragen (Suchwörter)
- den Zeitpunkt der Suche
- die angesteuerten Treffer (auch Werbung)
- Verweildauer auf den Treffern


Google- Geschäftsmodell

- Kontextbezogene Werbung zu den Suchwörtern
- Personenbezogene Werbung über angelegte Benutzerprofile
- Verkauf von persönlichen Daten und Nutzerprofilen an Firmen für ihre personenbezogene Werbung

Was macht geschicktes und professionelles Suchen aus?



Anfragetypen

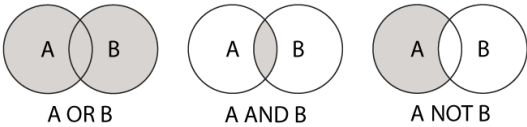


- Gezielte Suche nach einem bestimmten Faktum oder Dokument mit klar definiertem Ergebnis.
Optimal: Der erste Treffer ist richtig
- Recherche zu einem Thema nach bisher unbekanntem Informationen ohne klares Ende.
Optimal: Viele relevante Treffer

Formulieren von Suchanfragen

- Passende Suchwörter
 - nicht zu breite Bedeutung
 - verschiedene Schreibweisen beachten
 - Synonyme beachten
 - Phrasensuche in Anführungszeichen bei Zitaten
- Einengung der Suchanfrage
 - Boolesche Operatoren kombinieren Suchwörter
 - Erweiterte Suche mit Befehlen: Filetyp, Sprache


Suchwort-Verknüpfungen mit Booleschen Operatoren



A OR B	Hund OR Katze	122.000.000 Treffer
A AND B	Hund (AND) Katze	5.800.000 Treffer
A NOT B	Hund NOT Katze	90.300.000 Treffer

(Karotte OR Möhre) (Salat Nüsse) -Koriander

Such-Tipps



- Auf deutscher oder englischer Oberfläche der Suchmaschine anfragen, z.B. <https://www.google.fr>
- Gleich eine ausländische Suchmaschine verwenden, z. B. Bing
- Portale und Verzeichnisse sind Linksammlungen zu einem speziellen Thema.
 - Qualitätsprüfung durch ein Redakteursteam
 - nur themenrelevante Treffer
 - Verzeichnis der Verzeichnisse: www.suchlexikon.de

Metasuchmaschinen

- Haben keinen eigenen Index über Crawler angelegt.
- Suchanfragen werden an mehrere andere Suchmaschinen weitergeleitet.
- Treffer aus mehreren Trefferlisten werden neu gerankt.
- Metasuchmaschinen ergeben mehr Treffer, aber nicht unbedingt relevantere Treffer!

Welche Alternativen zu Google gibt es?



Suchmaschinen



DuckDuckGo

Metasuchmaschinen



Spezialsuchmaschinen

Beschränkung der Suche auf bestimmte Inhalte bzw. Dateitypen

- Suche nach Nachrichten
 - <https://news.google.de> (700 deutsche Quellen)
 - <https://de.nachrichten.yahoo.com>
- Suche nach wissenschaftlicher Literatur
 - <https://scholar.google.de>
- Jobsuchmaschinen
- Produktsuchmaschinen, Shopping

Spezialsuchmaschinen

- Suche nach Videos:
 - <http://www.google.de/videohp>
- Suche nach Personen
 - <http://www.yourtraces.com>
- Suche nach Zoologischen Gärten und Tierparks in Deutschland:
 - <http://zoogl.de>
- Suche nach Spezialsuchmaschinen
 - www.suchlexikon.de

Wie kann man die Seriosität der Treffer beurteilen?



Bewertung der Treffer

- **Glaubwürdigkeit**
 - Die Autoren von Texten sind genannt: Experten?
 - seriöse Institution, z.B. Hochschule, Verlag, Wissenschaftliche Gesellschaft, Qualitätspresse
 - Verlinkung: Auf welche Sites wird weiter verlinkt?
 - Vorsicht bei kommerziellen und gesponserten Sites
- **Aktualität**
 - Publikationsdatum und letzte Änderung (last update)
- **Bestätigung**
 - Angabe von Quellen, Belege für Behauptungen
 - Impressum, Kontaktinformation

Ausblick: semantische Suchmaschinen



- Faktensuchmaschine
- Suchanfragen sind frei als Fragen formulierbar.
- Es gibt keine Linkliste, sondern es werden direkt Antworten gegeben.

Buchtipp

