

Handout was können Suchmaschinen

Funktionsweise und Arten von Suchmaschinen

Suchmaschinen

Suchmaschinen eignen sich besonders gut, wenn Sie aktuellere Informationen oder Fakten suchen. Suchmaschinen sind im Prinzip nichts anderes als automatisch arbeitende Programme (Robots oder Crawler genannt). Diese durchsuchen das Internet Tag und Nacht nach neuen oder veränderten Internetseiten und speichern den kompletten Text der Internetseiten - Wort für Wort, Zeichen für Zeichen - in riesigen Datenbanken ab. Die Crawler starten meist bei einer bekannten Internetadresse und besuchen von dort jede Seite, die über einen Link erreichbar ist. Eine Qualitätsprüfung findet dabei nicht statt, allenfalls eine automatische Filterung, z.B. werden nicht alle Dateiformate gespeichert. Den Prozess der Speicherung nennt man Indexierung, die Datenbank der Suchmaschine daher auch den Index. Durch Eingabe von Suchbegriffen in ein Eingabefeld können Sie den Index durchsuchen.

Ein Crawler kann einige Milliarden Seiten pro Tag indexieren und die Indexe der größten Suchmaschinen umfassen viele Milliarden Internetseiten mit Billionen von Textzeichen. Da die Crawler nur Internetseiten indexieren können, die über einen Link erreichbar sind, entgehen ihnen aber viele Seiten - eben alle, die nicht über einen Link erreichbar sind oder die in irgendeiner Weise gegen Zugriffe geschützt sind. Zudem dringen Suchmaschinen nicht in die untersten Hierarchien von grösseren Seiten vor und haben nur ein bestimmtes Zeitbudget für die Indexierung einer Seite. Zudem können Suchmaschinen auch nicht alle Fileformate indexieren oder die Indexierung dauert länger. Wie beispielsweise auch bei JavaScript der Fall.

Algorithmus

Suchmaschinen erstellen ein Ranking der Indexierten Seiten zur Beantwortung einer Suchanfrage. Natürlich werden nicht alle Seiten im Ranking angezeigt. Suchmaschinen zeigen normalerweise weniger als 1000 Resultate an. Je nach Suchmaschine einiges weniger. Das Ranking wird durch einen Algorithmus erstellt. Dieser berücksichtigt dabei viele Faktoren. Schätzungsweise sollen es bei Google über 200 sein. Wobei der Algorithmus bei Google stetig weiterentwickelt wird. Sicher ist das Google die eingeben Suchanfrage, die Relevanz, Nutzerfreundlichkeit und Nützlichkeit von Seiten, die Verlässlichkeit von Quellen sowie den Standort und die Einstellungen berücksichtigt. Wobei die Gewichtung der Faktoren je nach Suchanfrage unterschiedlich ausfällt. Zum Erreichen eines möglichst hohen Rankings in der Google Suche hat sich eine ganze Search Engine Optimisation (SEO) Industrie gebildet.

Spezielsuchmaschinen

Spezielsuchmaschinen funktionieren im Prinzip genauso wie Suchmaschinen, allerdings werden von vornherein nur ganz bestimmte Internetseiten besucht und deren Inhalt indexiert. Dies können z.B. folgende Arten von Internetseiten sein:

- wissenschaftliche Internetseiten
- besonders aktuelle Internetseiten, z.B. Nachrichten
- Internetseiten zu einer Region, einer Sprache oder einem Land
- spezielle Dateiformate, z.B. Grafiken, Videos oder Musik
- Weblogs und andere Social-Media-Kanäle
- Adressen, Telefonnummern, Firmen

Dieser Bereich wird dann häufig vollständiger indexiert, als von einer allgemeinen Suchmaschine. Spezialsuchmaschinen bieten häufig auch mehr Suchmöglichkeiten an, die auf diesen Bereich optimiert sind. In einer Spezialsuchmaschine für Filme gibt es z.B. die Möglichkeit direkt nach Filmtiteln, Regisseuren, Schauspielern oder Charakteren zu suchen. Spezialsuchmaschinen eignen sich daher besonders gut, wenn Sie nur einen bestimmten Bereich im Internet durchsuchen wollen.

Metasuchmaschinen

Metasuchmaschinen funktionieren auf den ersten Blick wie "normale" Suchmaschinen. Sie geben etwas in ein Suchfeld ein und erhalten eine Trefferliste. Metasuchmaschinen besitzen jedoch keinen eigenen Index, sondern machen sich die Vorarbeit von anderen Diensten zunutze und leiten Ihre Suchanfragen an verschiedene Suchdienste (dies können Suchmaschinen, Spezialsuchmaschinen, Linksammlungen oder Nachschlagewerke sein - zum Teil sogar andere Metasuchmaschinen) und stellen die Ergebnisse in einer Trefferliste zusammen. Das läuft wie folgt ab:

Sie geben eine Suchanfrage bei einer Metasuchmaschine ein. Die Metasuchmaschine gibt Ihre Anfrage an verschiedene Suchdienste weiter. Die einzelnen Suchdienste suchen in ihren Datenbanken und liefern der Metasuchmaschine die Ergebnisse (i.d.R. aber nur die ersten 10 Treffer). Die Ergebnisse werden von der Metasuchmaschine gesammelt, zusammengestellt und in einer gemeinsamen Trefferliste angezeigt.

Metasuchmaschinen ermöglichen so zwar die gleichzeitige Suche in vielen verschiedenen Suchdiensten, dafür müssen Sie aber oft Einschränkungen bei Suchfunktionen und auch bei der Trefferqualität hinnehmen. Häufig werden die Ergebnisse auch noch mit Werbung "angereichert".

Quelle: <https://www.ub.uni-bielefeld.de/ub/learn/tutorials/websearch/overview.xml> und <https://www.ub.uni-bielefeld.de/ub/learn/tutorials/websearch/invisibleweb.xml> mit Ergänzungen von Fabian

Konsequenzen

Die Funktionsweise von Suchmaschinen, allen voran Google, führt zu einigen Problemen. Wichtig ist zu erwähnen, dass Suchmaschinen nicht alles finden. **Den auch Google indexiert nicht einmal annähernd das gesamte Internet. Inhalte von Datenbanken werden beispielsweise nicht indexiert.** Deswegen ist die Recherche in Bibliothekskatalogen, Datenbanken und anderen Ressourcen, die auch im Handout Rechercheeinführung aufgeführt sind, wichtig.

Wie bereits erwähnt ist eine ganze Industrie entstanden, die sich auf das Optimieren von Webseiten für den Suchalgorithmus spezialisiert haben. Konsequenterweise bedeutet, dass das Internet auf den Google-Suchalgorithmus zugeschnitten ist und nicht auf Internetbenutzer. Zudem wird viel AI verwendet, um Seiten für den Algorithmus zu kreieren, was zu einer Abnahme der Qualität des Internets führt.

Die Qualität der Ergebnisse auf der ersten Google-Suchseite ist mittlerweile fragwürdig geworden. Zudem sind die echten Suchergebnisse immer weiter unten auf der Ergebnisliste. Denn Google bewirbt zuerst einmal seine eigenen Angebote, dann die von Sponsoren und schließlich werden die Suchergebnisse auch noch durch KI zusammengefasst, bevor die eigentlichen Resultate angezeigt werden. **Die Qualitätsprüfung dieser Zusammenfassungen wird auf Google allerdings verunmöglicht. Bitte schaut euch die Originalquelle an, um festzustellen ob, die gelieferten Informationen vertrauenswürdig sind.**

In diesem Sinne läuft momentan eine interessante Diskussion um den Einfluss von KI auf das Internet. Es wird befürchtet, dass der von Menschen gemachte Content von einem riesigen Mengen an KI-

Content überflutet wird und von Menschen erstellter Content bald schwer zu finden sein wird. Da KIs durch Daten aus dem Internet trainiert werden, könnte das schliesslich auch die Qualität der KI beeinflussen. Falls euch das interessiert, gibt es mehr Informationen [hier](#), [hier](#) oder [hier](#) (Die Seiten sind auf Englisch).

Erweiterte Suche mit Google

Google und andere Suchmaschinen bieten auch die Möglichkeit einer erweiterten Suche. Leider versteckt Google diese immer besser. Die Suche ist unter https://www.google.de/advanced_search abrufbar oder man kann danach Googlen.

Neben der erweiterten Suche bietet Google auch noch spezifische Suchseiten an. Da wäre einerseits [Google Scholar](#) zur Suche von wissenschaftlichen Artikeln und andererseits [Google Books](#) zur Suche von und in Büchern. Die digitalisierten Bücher sind teilweise auch im Volltext verfügbar.

Google Suchoperatoren

Suchmaschinen unterstützen auch Suchoperatoren, um die Suche zu verbessern. Google unterstützte einmal 48 Suchoperatoren, wobei der Konzern immer mehr Operatoren abschafft. Hier ist eine Tabelle aller noch mehr oder weniger aktiven Suchoperatoren sortiert nach absteigender Relevanz:

Operator	Funktion	Beispiel
“ “	Exakte Suche von dem Wort in den Anführungszeichen. Verwandte Ergebnisse und Synonyme werden ausgeschlossen.	“walt disney“
-	Schliesst einen Begriff aus.	Darth -Vader
*	Ein Platzhalter, um unbekannte Begriffe zu ersetzen.	Brot mit *
Site:	Reduziert die suche auf eine Domain. Besonders für Foren geeignet	Site: reddit.com
_	Ein Platzhalter für Autocomplete. Erzeugt eine Liste in den Suchbegriffen	darth_ star wars (so würden alle Sith Lords durch Google Autocomplete aufgelistet)
OR oder	Zeigt Resultate zum einen, zum anderen oder zu beiden Begriffen an.	Ketchup Mayo
()	Gruppierung von Wortfolgen.	(Ewan McGregor OR Alec Guinness) Obi Wan Kenobi
cache:	Liefert den letzten Cache einer Seite.	cache: disney.com
source:	Reduktion auf eine Quelle in Google News	source: guardian.co.uk
AROUND(x)	Sucht Seiten die zwei Begriffe innerhalb der definierten Anzahl Wörter enthalten.	Jochen AROUND(3) Verräter
intitle:	Zeigt nur Seiten an, die ein bestimmtes Wort im Titel enthalten.	intitle: liste
inurl:	Zeigt nur Seiten an, die ein bestimmtes Wort in der URL enthalten.	inurl: ressourcen
allintitle:	Zeigt nur Seiten an, die alle Worte im Titel enthalten.	allintitle: top liste
allinurl:	Zeigt nur Seiten an, die alle Worte in der URL enthalten.	allinurl: liste dichter
imagesize:	Bestimmung der Grösse in Pixeln in der Bildersuche.	imagesize: 1920 x 1080
filetype:	Zeigt Seiten mit einem bestimmten Dateityp an.	filetype: pdf

ext:	Zeigt Seiten mit einer bestimmten Dateierdung an.	ext: pdf
before:	Zeigt Seiten mit einem Aktualisierungs- oder Veröffentlichungsdatum vor einem festgelegten Tag oder Jahr an.	Marvel before: 2009-12-31
after:	Zeigt Seiten mit einem Aktualisierungs- oder Veröffentlichungsdatum nach einem festgelegten Tag oder Jahr an.	Marvel after: 2009-12-31
location:	Artikel zu einem bestimmten Ort in Google News	location: vaduz
define:	Erzwingt die Darstellung eines Wörterbucheintrags	define: greif
In	Umrechner	100 m in inches
intext:	Zeigt nur Seiten an, die ein bestimmtes Wort im Text enthalten.	intext: Lyrik
stocks:	Zeigt Aktieninformationen an	stocks: disney
weather:	Zeigt Wetterinformationen an	weather: balzers
map:	Erzwingt eine Karte.	map: vaduz
movie:	Informationen zu einem Film	movie: daredevil
AND	Suche muss beide Begriffe enthalten.	Beer AND pong

Die Suchoperatoren können nach Belieben miteinander kombiniert werden.

Alternativen zu Google

Suchmaschine	Erläuterung
Wolfram Alpha	Wolfram Alpha ist eine semantische Suchmaschine. Die Ergebnisse sind synthetische Antworten, von einer simplen Summenbildung bis zu Eigenwerten von Differenzialgleichungen oder Umlaufbahnen von Planeten.
Swurl	Alternative Meta-Suchmaschine. Für suchen in Google und Social Media zur gleichen Zeit.
Yep	Suchmaschine mit Schutz der Privatsphäre und Revenue Sharing. Höchste Updatefrequenz nach Google.
Carrot 2	Die Suchergebnisse werden in Themenbereiche geordnet.
Brave	US-Suchmaschine ohne Tracking und Werbung
Duckduckgo	Die Suchmaschine DuckDuckGo speichert keine Suchhistorie und funktioniert ohne Filterbubble.
Startpage	Proxy-Metasuchmaschine zum Absuchen von Google oder Bing ohne spuren zu hinterlassen.
Suchmaschinen-Datenbank	Suchmaschine zur Suche der passenden (Spezial)suchmaschine.
Crowd View	Suchmaschine für Foren
Feedle	Suchmaschine für Podcasts
Metager	Die vom Nutzer eingegebenen Suchwörter werden gleichzeitig in ca. 20 bis 30 Suchmaschinen gesucht. MetaGer wird betrieben vom gemeinnützigen Verein SUMA-EV in Kooperation mit der Leibniz Universität Hannover, speichert keine Nutzerdaten und ermöglicht es die Ergebnis-Links anonym zu öffnen.
Consensus	Consensus ist eine wissenschaftliche Suchmaschine, die mit KI arbeitet.