

Spider Suchmaschinen

Spider Suchmaschinen basieren auf einem Suchroboter und bestehen im Wesentlichen aus drei Teilen:

- Der erste Teil ist der Informationssammler, "Robot", "Spider", oder "Crawler" genannt, der im WWW Seiten vollautomatisch absurft. Einer oder mehrere dieser Exemplare sorgen dafür, dass ein riesiger Datenberg an die "Verwaltung", den Index geschickt wird.
- Der zweite Teil ist die Indizierungssoftware, die die Daten strukturiert und durchsuchbar macht.
- Der dritte Teil ist eine Software zum Auswerten von Suchanfragen. Diese verknüpft logische Operatoren, und schickt die Anfrage an den Datenbank-Server, um von dort aus die Ergebnisse zu präsentieren. Hierbei wird meist auch der Ort des Vorkommens im Dokument besonders berücksichtigt. Wenn das Wort im Titel oder in den Meta-Tags des HTML-Dokuments vorkommt, wird es höher gewichtet als im Body-Text.

Doch was sind eigentlich Meta-Tags? Meta-Tags stehen im Header (Kopf) des HTML-Dokuments und werden vom Browser nicht angezeigt. Sie definieren u.a. die Keywords (Schlüsselwörter), nach denen das Dokument in der Suchmaschine gefunden werden soll, und eine kurze Zusammenfassung des Seiteninhalts.

Bei der Suche in einer Spider Suchmaschine muss anders als bei der Suche in einem Webkatalog mehr auf der Ebene des Computers gedacht werden und weniger wie ein Redakteur. Die Strukturierung eines Dokumentes und die logische Verknüpfung von Begriffen tritt in den Vordergrund. In den Hintergrund gerät der sachliche Gesamtinhalt und die Relevanz einzelner Worte für diesen Inhalt. Diese Relevanz muss durch eine geschickte Formulierung der Anfrage bestimmt werden. Zu diesem Zweck sollte man sich möglichst der erweiterten Suchoptionen bedienen. Hier sind die Verknüpfungsmöglichkeiten viel größer.

Zum Beispiel ist es vollkommen unsinnig, nach einem häufigen Begriff, wie z. B. "Wasser" zu suchen. Die Trefferliste würde in die Millionen gehen. Vielmehr muss das Umfeld definiert werden und man muss sich über den Zusammenhang klar werden, in dem der Begriff gesucht werden soll, z.B. Energiegewinnung oder Trockenheit.

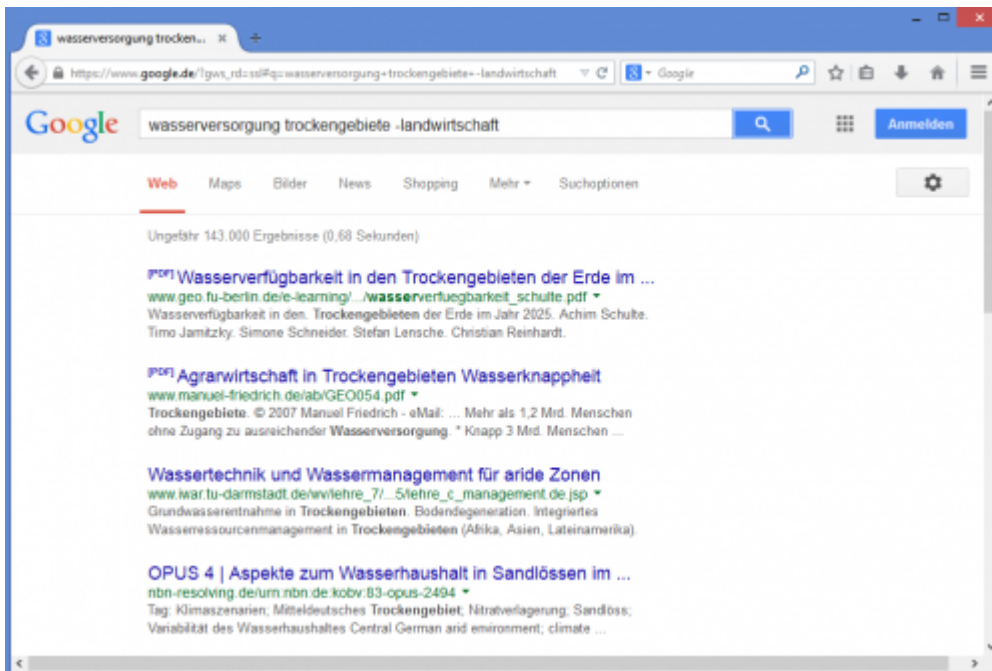
Weiter sollte man sich überlegen, welche Begriffe noch in Zusammenhang mit dem Suchwort stehen und welche explizit ausgeschlossen werden können. Beim Beispiel "Wasser" lassen sich Begriffe wie "Energie", "Schifffahrt" und "Medizin" ausschließen. Das Suchgebiet ist immer noch ziemlich groß. Wenn man z.B. wissen möchte, welche Rolle die künstliche Wasserversorgung in der Landwirtschaft spielt, könnten Begriffe wie "Bewässerung", "Versorgung", "Pumpen", "Rohre", "Trockenheit" oder der Terminus "künstliche Wasserversorgung" die Zahl der Dokumente besser spezifizieren. Man sollte aber nicht mit zu vielen UND-Begriffen beginnen, da eventuell die Datenbasis nicht ausreichend genug ist und man dann eventuell gar keine Treffer erhält. Ein schrittweises Eingrenzen bei zu vielen Treffern bietet sich an.

Hier sind einige häufig zu findende Eingabemöglichkeiten:

- Ein Pluszeichen (+) oder AND verknüpft das nachfolgende Wort mit dem vorherigen. Beide Begriffe müssen im Ergebnisdokument vorkommen. Mehrere Begriffe lassen sich so zusammenfassen. Bei vielen Suchmaschinen hat ein Leerzeichen zwischen zwei Suchbegriffen die gleiche Wirkung.
- Ein Minuszeichen (-) oder NOT schließt das nachfolgende Wort aus. Das Ergebnis darf das Wort nicht enthalten.
- Mehrere Worte lassen sich mit Anführungszeichen zu einer Phrase verbinden. Diese werden dann als ein einzelner Begriff behandelt: z.B.: "Albert Einstein" oder "französische Küche".

Die Syntax der Suchserver bei der Eingabe, die Möglichkeiten

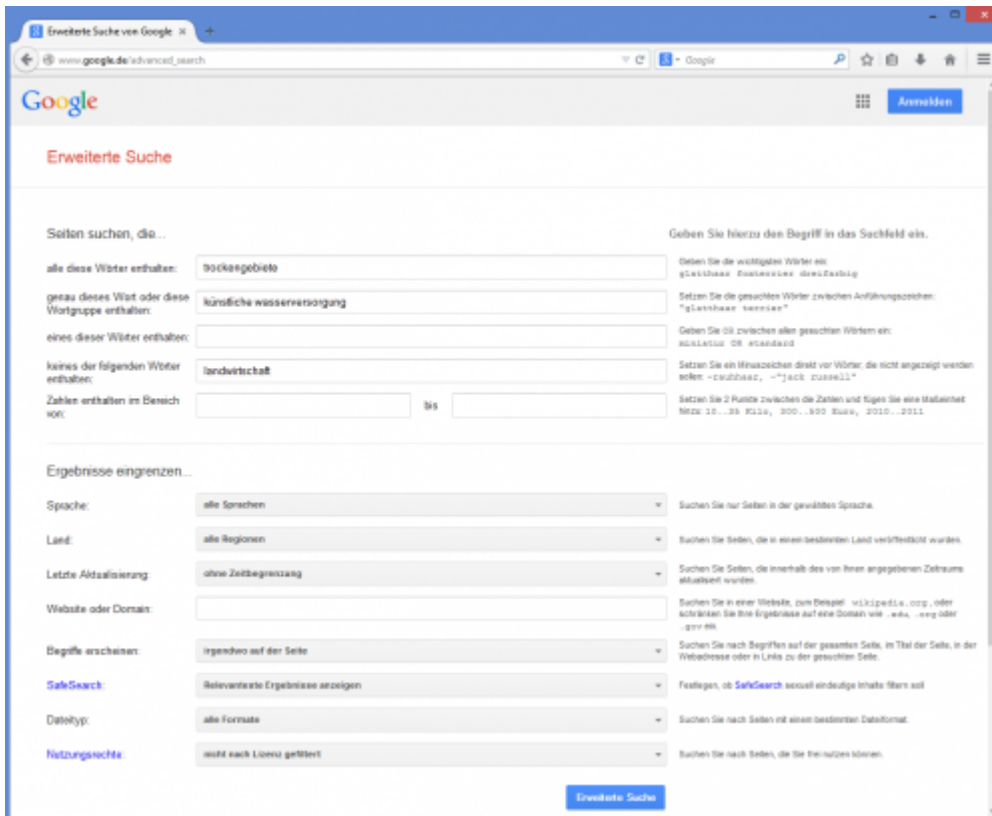
und der Komfort unterscheiden sich zum Teil erheblich voneinander. Genauere Informationen hierzu gibt es in der Suchfibel.



Entsprechend dem oben genannten Beispiel soll an die Suchmaschine von Google eine Suchanfrage nach dem Begriff "Wasser" gestellt werden. Hier zeigt sich, dass durch Einschränkung des Suchbegriffs anhand der Suchanfrage:

wasserversorgung trockengebiete -landwirtschaft

die Trefferzahl mit 143.000 gefundenen Seiten schon stärker eingegrenzt wird, aber für eine Verwertung noch zu hoch ist. Über die erweiterte Suche in Google wird in unserem Beispiel der Begriff wasserversorgung durch die genaue Wortgruppe "künstliche Wasserversorgung" ersetzt. Hierbei werden die Treffer schon weiter eingegrenzt.



Bei einer Suche nach **“künstliche Wasserversorgung” - landwirtschaft** werden 501 Treffer gefunden und bei einer Suche nach **trockengebiete “künstliche Wasserversorgung”** sogar nur 1 Treffer. Die Suche nach **trockengebiete “künstliche Wasserversorgung” -landwirtschaft** würde allerdings die Suche schon so weit eingrenzen, dass überhaupt kein Treffer mehr gefunden wird.

Eine andere Suchstrategie hat hier vielleicht mehr Erfolg. Die erweiterte Suche von Google ermöglicht auch die Verwendung von Suchfeldern. Durch die Suchanfrage **allintext: trockengebiete “künstliche wasserversorgung” -landwirtschaft** wird bei Eingabe der Position des Vorkommens der Begriffe innerhalb der Antwortseiten **“im Text der Seite”**, eine Trefferzahl von 576 erzielt.

An diesem Beispiel ist zu erkennen, dass die Wahl der richtigen Suchstrategie bei Spider Suchmaschinen nicht so einfach ist und man von Fall zu Fall immer wieder neu überdenken sollte.

Die wichtigsten Spider Suchmaschinen sind:

- Google,
- Yahoo und
- Bing.