**Der Index**

Der Indexer erzeugt aus einer gesammelten Seite nach einem bestimmten Algorithmus einen Index. Beispielhaft wollen wir das an folgendem Beispiel ausführen. Die folgende Seite wurde unter <http://www.mondladung.ch> vom Crawler gefunden.

Am 20. Juli 1969 landeten die Amerikaner erstmals auf dem Mond. Neil Armstrong berührte den Mond als Erster und verkündete: „Ein kleiner Schritt für einen Menschen, aber ein Riesenschritt für die Menschheit.“

Zuerst wird festgestellt, um welche Sprache es sich handelt. Da die Seite in Deutsch ist, wird nun eine Buchstabenumwandlung gemacht. ä, ö, ü wird zu ae, oe, ue.

Am 20. Juli 1969 landeten die Amerikaner erstmals auf dem Mond. Neil Armstrong berührte den Mond als Erster und verkündete: „Ein kleiner Schritt für einen Menschen, aber ein Riesenschritt für die Menschheit.“

Anschliessend wird die Wortextraktion durchgeführt, d.h. es werden die Satzzeichen weggelassen.

Am 20. Juli 1969 landeten die Amerikaner erstmals auf dem Mond. Neil Armstrong berührte den Mond als Erster und verkündete: „Ein kleiner Schritt für einen Menschen, aber ein Riesenschritt für die Menschheit.“

Nun werden alle Stoppwörter eliminiert. Stoppwörter sind Begriffe, die nichts oder nur sehr wenig zur Beschreibung des Inhalts beitragen. Beispiel: am, die, auf,...

Am 20. Juli 1969 landeten die Amerikaner erstmals auf dem Mond. Neil Armstrong berührte den Mond als Erster und verkündete: „Ein kleiner Schritt für einen Menschen, aber ein Riesenschritt für die Menschheit.“

In der nächsten Phase werden die Wörter zerlegt und normalisiert. So wird etwa aus „Riesenschritt“ „riesig“ und „Schritt“. Bei der Normalisierung wird lediglich der Wortstamm, ohne Endungen und verschiedene Schreibweisen, übernommen. Etwa statt „beruehrte“ „beruehr“. Zudem werden alle Groß- in Kleinbuchstaben umgewandelt.

Am 20. Juli 1969 landeten die Amerikaner erstmals auf dem Mond. Neil Armstrong berührte den Mond als Erster und verkündete: „Ein kleiner Schritt für einen Menschen, aber ein Riesenschritt für die Menschheit.“

Die so behandelte Seite wird nun iniziert. Dazu werden die einzelnen Begriffe im Dokument gezählt und in einer Tabelle aufgelistet.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Begriff | Häufigkeit | Positionen | ... | ... |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| ... | ... | ... |  |  |

Der Suchdienst weiss nun, welche Begriffe sich auf der Seite [www.mondlandung.ch](http://www.mondlandung.ch) befinden. Wird nun eine Suchanfrage wie beispielsweise „mondlandung erste“ gestellt, so erkennt der Suchdienst, nachdem er die Suchanfrage ebenfalls in „mond“ „landung“ „erst“ zerlegt hat, dass die Seite [www.mondlandung.ch](http://www.mondlandung.ch) dieser Anfrage entspricht und liefert sie als Treffer zurück. Wo in der Rangliste die Seite dann erscheint, hängt von den Rangierungsprinzipien ab.