

Crawler

Ein Crawler, abgeleitet von Webcrawler, ist ein **automatisiertes Programm**, welches das World Wide Web nach Seiten durchsucht, diese auf den Inhalt analysiert und anschließend auflistet. Der **Crawler erstellt** durch seine Arbeit sozusagen das **Inhaltsverzeichnis für die Suchmaschinen**, in dem alle Inhalte der Webseiten gespeichert sind, regelmäßig automatisch aktualisiert werden und für die Suchmaschinen im Index (Datenbank) hinterlegt sind. Erst durch die unzähligen Crawler ist also die Indexierung der meisten Webseiten in den Suchmaschinenergebnissen überhaupt möglich.

Die meisten **Crawler gelangen über die Hyperlinks auf bestehenden Webseiten zu anderen Webinhalten**, sortieren diese neu und indizieren diese dann gegebenenfalls neu für die Suchmaschinen. Je nach Größe der Webseite wird diese entsprechend häufig von den Crawlern auf aktualisierte Inhalte erneut durchsucht, welche diesen dann speichern und wieder zurück an den Index der jeweiligen Suchmaschine übermitteln. Wichtig ist aber, dass die Webseite, die gecrawlt, also durchsucht werden soll, in den Einstellungen der Robots.txt-Datei den Webcrawlern den Zugriff gewährt.

Will man den gesamten Inhalt einer Webseite für den Webcrawler zur Verfügung stellen, empfiehlt es sich, ein XML-Sitemap zur Verfügung zu stellen, welche alle Unterseiten einer Domain auflistet. Synonym zu Crawlern wird im Online-Marketing auch oft von *Spidern* oder *Robots* gesprochen.