Robots.txt

Was ist die robots.txt Datei?

Eine robots.txt Datei ist eine Textdatei, in der man hinterlegen kann, welche Verzeichnisse von den Suchmaschinen gelesen werden dürfen und welche nicht. Die robots.txt Datei ist sehr wichtig für die Crawler, diese suchen als Erstes die Datei und lesen diese aus. Wie eine Anleitung was Sie tun dürfen und was nicht.

In der robots.txt wird genau hinterlegt, welcher Crawler bestimmte Verzeichnisse, Unterverzeichnisse oder nur einzelne Dateien nicht durchsuchen darf und für ihn gesperrt sind. Sie können also hinterlegen, dass der googlebot bestimmte Seiten durchsuchen darf, aber der bingbot nicht, oder andersherum.

Wohin mit der robots.txt Datei?

Damit die Suchmaschinen die robots.txt finden, muss sich diese im Hauptverzeichnis der Domain befinden. Sollten Sie die Datei woanders speichern, wird diese von den Suchmaschinen nicht gefunden und beachtet.

Es kann pro Hauptdomain nur eine robots.txt Datei geben.

Aufbau und Inhalt einer robots.txt Datei

Die robots.txt Datei besteht aus zwei Elementen, die immer zusammengehören. Erst spricht man den User-Agenten an und nennt seinen Namen. Darunter kommt der Befehl mit dem Namen des Verzeichnisses, das gelesen werden soll oder nicht.

Sie können die Sitemap.xml Datei in der robots.txt Datei hinterlegen und somit sicher sein, dass der Crawler diese überhaupt aufruft.

Als Erstes nennt man den Befehl, der den Bot anspricht. **User-agent:**

Dahinter kann man eine genaue Bezeichnung einfügen oder mit * kennzeichnen, dass man alle Bots ansprechen möchte.

SEO-Küche Internet Marketing GmbH & Co. KG
Fraunhoferstr. 6, 83059 Kolbermoor
Telefon 08031 / 2575-100
Telefax 08031 / 2575-101
E-Mail: info@seo-kueche.de

SEO-Küche Internet Marketing GmbH & Co. KG,
HRA 11167 AG Traunstein
pers. Haftende Gesellschafterin:
SEO-Küche Verwaltungs GmbH, Kolbermoor, HRB 22414 AG Traunstein
Geschäftsführer: Christian Brunnenmayer, Patrick Keller, Oliver Lindner
Ust-IdNr.: DE 286 985 708, Steuer Nr. 156/174/08500

HypoVereinsbank
IBAN DE45 700202700015260147
BIC HYVEDEMMXXX

Ein Befehl wie disallow**:** schließt alle betroffenen Dateien aus.

Der Befehl wie Allow: / schließt alle betroffenen Dateien ein.

Hier ein Beispiel einer robots.txt Datei:

Beispiel 1:

User-agent: seobot

Disallow: /noseobot/

Das heißt der User-Agent mit dem Namen „seobot“ soll den Ordner http://www.test.de/nosebot/ inklusive Unterverzeichnisse nicht crawlen.

Beispiel 2:

User-agent: *

Allow: /

Im Beispiel zwei wird gezeigt, dass alle User-Agents auf die gesamte Website zugreifen sollen. Diese Regel ist unnötig, da die Crawler automatisch alles crawlen, wenn kein anderer Befehl sie davon abhält.

Beispiel3:

User-agent: seobot

Disallow: /verzeichniss1/

Disallow: /verzeichniss6/

Hier haben wir dem seobot gesagt, dass das Verzeichnis 1 und 6 für ihn gesperrt sind und er diese nicht durchsuchen darf.

Hier eine Auswahl der wichtigsten User-Agenten Bezeichnungen:

Crawler	User-agent
Google	Googlebot
Bing	Bingbot
Yahoo	Slurp
MSN	Msnbot

Kann ich der robots.txt Datei blind vertrauen?

Die robots.txt Datei ist nur eine Hilfe für die Crawler, es ist nicht sicher garantiert, dass die Websiten nicht gecrawlt werden. Auch ein Schutz vor Zugriffen anderer schützt die robots.txt Datei nicht, hier sollte man immer mit einem Passwortschutz des Webservers arbeiten. Goolge und Bing geben beide an, dass Sie die robots.txt Datei beachten, sie sind aber nicht verpflichtet dazu.

Wie rufe ich die robots.txt Datei auf?

Sie können die Datei in Ihrem Browser ganz einfach aufrufen. Tragen Sie oben in der URL Leiste einfach Ihre Domain ein und dahinter /robots.txt.

Beispiel: www.ihrdomainname.de/robots.txt

Die Datei soll nur mit der Hauptdomain aufgerufen werden können. Sollten Sie die Datei wie folgt aufrufen: www.ihredomain.de/verzeichnis/robots.txt, sollte hier ein 404 Fehler aufkommen. Wenn Sie mit dieser Eingabe aber auf die Startseite von Ihrer Website gelangen, überprüfen Sie Ihre Weiterleitungen. Dies führt zwar zu keinem Fehler, aber es ist nicht richtig. Eine Seite, die nicht erreichbar ist, sollte auch als solche ausgegeben werden.

SEO-Küche Internet Marketing GmbH & Co. KG
Fraunhoferstr. 6, 83059 Kolbermoor
Telefon 08031 / 2575-100
Telefax 08031 / 2575-101
E-Mail: info@seo-kueche.de

SEO-Küche Internet Marketing GmbH & Co. KG,
HRA 11167 AG Traunstein
pers. Haftende Gesellschafterin:
SEO-Küche Verwaltungs GmbH, Kolbermoor, HRB 22414 AG Traunstein
Geschäftsführer: Christian Brunnenmayer, Patrick Keller, Oliver Lindner
Ust-IdNr.: DE 286 985 708, Steuer Nr. 156/174/08500

HypoVereinsbank
IBAN DE45 700202700015260147
BIC HYVEDEMMXXX