

Robot

Was ist ein Robot?

Ein **Robot** ist ein Programm, das für die [Suchmaschinen](#) selbstständig das Internet nach neuen Inhalten durchsucht und diese erfasst, bewertet und indiziert. Statt Robots wird häufig auch die Kurzform "Bot" genutzt. Alternative Bezeichnungen sind **Spider** oder **Crawler**. Heutzutage bestehen etwa 40 Prozent des weltweiten Webtraffics aus Robots.

Funktionsweise eines Robots

Der Robot durchsucht automatisiert Dokumente im Web und folgt dabei auch **externen und internen Links**. Dadurch kann er beispielsweise feststellen, welche Linkbeziehungen zwischen welchen [Webseiten](#) bestehen. Bevor er "loslegt" benötigt der Robot natürlich eine genau definierte Aufgabe. Dadurch kann er bspw. Informationen einer Kategorie zuordnen. Nach und nach werden so alle Seiten erfasst und katalogisiert. Da der Robot seine Aufgabe ständig wiederholt, werden auch neue Informationen mit aufgenommen und indiziert.

Es gibt jedoch auch Informationen/Seiten, die dem Robot verwehrt bleiben. Das passiert z.B. bei Seiten, die nur durch einen Login oder über eine Suchmaske erreichbar sind.

Einsatzmöglichkeiten für Robots

In der Regel werden Robots dazu genutzt, Indizes zu erstellen, die für Suchmaschinen die Basis für die Einordnung von Suchergebnissen liefern. Dabei werden die Robots so programmiert, dass sie ganz spezifisch suchen, z.B. nach aktuellen Nachrichten oder nur bestimmten thematisch relevanten Links folgen.

Beispiele für **Robot-Typen**:

Data-Mining: Robots suchen gezielt nach Daten wie E-Mail-Adressen oder Telefonnummern.

Webanalyse: Robots sammeln u.a. Daten zu Seitenaufrufen, Seiteneinstiegen usw.

Produktrecherche für Preisvergleiche: Robots sammeln Produktdaten und Produktpreise, damit sie der User auf den Portalen vergleichen kann.

Robot lenken über die Robots.txt

Sie als Website-Betreiber können Robots auf Ihrer Seite mithilfe einer Datei lenken und so deren Verhalten auf Ihrer Webseite regeln. Diese Datei trägt den Namen "Robots.txt" und ist in der **Root Directory** (Stammverzeichnis) Ihrer Webseite untergebracht. Das ist die erste Datei, die von den Robots gelesen wird und mit der Sie die Programme steuern können. Mit genauen Anweisungen können Sie den Robot, genauer gesagt den [User Agent](#) des Robots so lenken, dass er nur ganz bestimmte Wege geht bzw. nur ganz bestimmte Seiten aufsucht. Seiten zu sperren ist z.B. bei einem Webseitenumbau sinnvoll, um dem Robot keine ungewollten/falschen Daten zu liefern.

Auch wenn Sie über die Robots.txt vorgeben, was der Robot nicht durchsuchen darf, können diese Seiten trotzdem von der Suchmaschine indexiert werden. Um das zu verhindern, nutzen Sie in den [Meta-Tags](#) den Noindex-Tag oder den [Canonical Tag](#).

Bsp. für eine Robots.txt:

User-agent: *

Disallow: /cgi-bin/

Disallow: /wp-admin/

Disallow: /trackback/

Disallow: /feed/

Disallow: /comments/

Disallow: /wp-content/uploads/

Disallow: */trackback/

Disallow: */feed/



Disallow: */comments/

Sitemap: <https://www.seo-kueche.de/sitemap.xml>

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