

Robots.txt

http://de.wikipedia.org/wiki/Robots_Exclusion_Standard

Die Suchmaschinen suche nach der Datei robots.txt. In dieser Datei ist festgelegt, welche Ordner oder Dateien die Suchmaschinen durchsuchen und indizieren darf.

Die Datei robots.txt liegt immer im ROOT-Verzeichnis

z.B.

<http://www.beispiel.at/robots.txt>

<http://test.beispiel.at/robots.txt>

.) Suchmaschinen dürfen alle Ordner und Dateien indizieren
(Mindestangabe)

Der Stern * ist ein jokerzeichen und bedeutet „Alle“

User-agent: *

Disallow:

.) Damit sperren wir den Webpace für alle Crawler.

User-agent: *

Disallow: /

.) Damit sperren wir den Webpace für den Googlebot.

User-agent: Googlebot

User-agent: MSNBot

Disallow: /

.) Damit sperren wir das Verzeichnis „kurs“

User-agent: *

Disallow: /kurs/

.) Damit sperren wir das Verzeichnis „kurs“ erlauben aber die Datei test.htm im Verzeichnis Kurs zu indizieren.

User-agent: *

Disallow: /kurs/

Disallow: /kurs/

.) Damit sperren wir alle Seiten die ein Fragezeichen in der URL enthalten.

User-agent: *

Disallow: /*?

.) Damit sperren wir alle Seiten die ein Fragezeichen in der URL enthalten, Dateien die ein fragezeichen am Schluß haben sollen hingegen indiziert werden.

User-agent: *

Disallow: /*?

Allow: /*?\$

.) Alle Dateien mit der Endung .doc sollen **NICHT** indiziert werden

User-agent: *

Disallow: /*.doc\$

.) Alle Verzeichnisse die mit beginnen werden nicht indiziert (z.B. /wifikurs/, /wifitraining/, etc).

User-agent: *

Disallow: /wifi*/

.) Die Sitemap liegt unter folgender Adresse (Google, Yahoo, MSN, ASK)

Sitemap: <http://www.beispiel.at/sitemap.xml>

Ein gesamtes Beispiel:

User-agent: MSNBot

Disallow: /

User-agent: Googlebot

User-agent: Scooter

Disallow:

User-agent: *

Disallow: /kurs/

Disallow: /test.html

Disallow: /css/

Sitemap: <http://www.beispiel.at/sitemap.xml>

Sitemap erstellen:

<http://www.sitemaps.org/de/protocol.php>