

Google und die Dateien die nicht in den Such-Index gehören

Es gibt Dateien die selbst für Suchmaschinen oder sogar insbesondere für sie relevant sind, die aber nichts im Such-Index zu suchen haben. So ist darauf zu achten, dass auch nur Dateien in den Such-Index geraten, die dort hin gehören also z.B. nicht die folgende Dateien:

- robots.txt
- sitemap.xml
- rss.rdf
- atom.xml

Beispiele wie Sie diese Dateien finden wenn Sie sich nicht sicher sind:

```
site:[your domain] robots
site:[your domain] sitemap
```

Für den Apache Webserver sollten in der .htaccess die folgenden Passagen eingefügt werden, damit die sitemap.xml und die robots.txt nicht im Google Index erscheinen:

```
# Robots noindex sitemap.xml
<IfModule mod_headers.c>
  <FilesMatch "sitemap\.xml$" >
    Header append X-Robots-Tag "noindex"
  </FilesMatch>
</IfModule>

# Robots noindex robots.txt
<IfModule mod_headers.c>
  <FilesMatch "robots\.txt$" >
    Header append X-Robots-Tag "noindex"
  </FilesMatch>
</IfModule>
```

Falls Sie direkt alle XML-, RDF- und TXT-Dateien aus dem Google Index verbannen wollen:

```
# Robots noindex *.xml,*.rdf,*.txt
<IfModule mod_headers.c>
  <FilesMatch "\.(xml|rdf|txt)" >
    Header append X-Robots-Tag "noindex";
  </FilesMatch>
</IfModule>
```

Sicherlich macht es Sinn bei dynamisch generiertem Inhalt, der nicht indiziert werden soll, im Header diese Werte direkt bei der Erzeugung zu setzen!

Autor: [Udo Schmal](#), veröffentlicht: 09.02.2013, letzte Änderung: 29.08.2024

© Copyright 2024 Udo Schmal