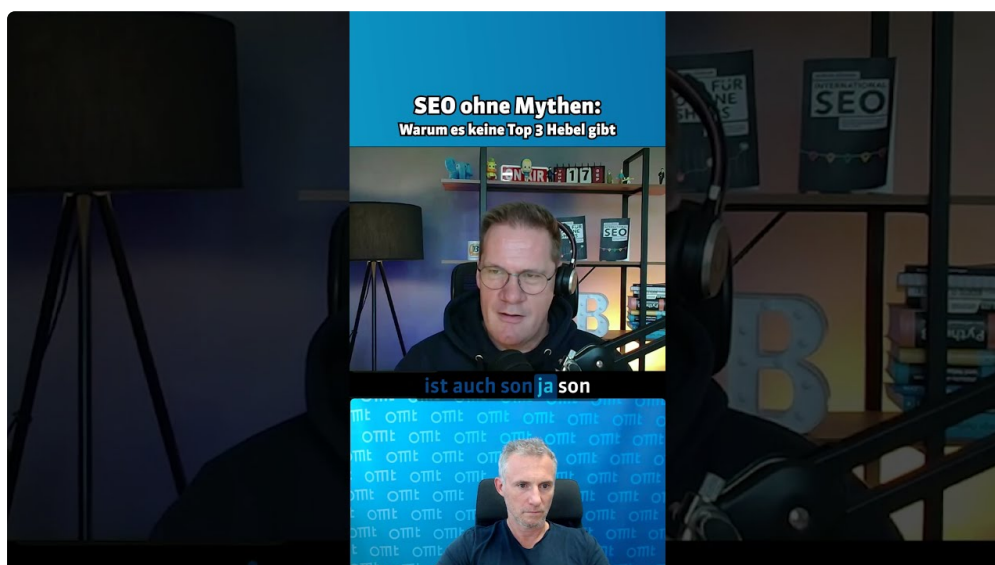


How to keep robots out of your web site

THE ROBOTS.TXT FILE You know that search engines have been created to help people find information quickly on the Internet, and the search engines acquire much of their information through robots (also known as spiders or crawlers), that look for web pages for them. The spiders or crawlers robots explore the web looking for and recording all kinds of information. They usually start with URL submitted by users, or from links they find on the web sites, the sitemap files or the top level of a site. Once the robot accesses the home page then recursively accesses all pages linked from that page. But the robot can also check out all the pages that can find on a particular server. After the robot finds a web page it works indexing the title, the keywords, the text, etc. But sometimes you might want to prevent search engines from indexing some of your web pages like news postings, and specially marked web pages (in example: affiliates pages), but whether individual robots comply to these conventions is pure voluntary.



ROBOTS EXCLUSION PROTOCOL So if you want robots to keep out from some of your web pages, you can ask robots to ignore the web pages that you don't want indexed, and to do that you can place a robots.txt file on the local root server of your web site.



In example if you have a directory called e-books and you want to ask robots to keep out of it, your [seo agency for startups punjab](#) robots.txt file should read: User-agent: * Disallow: e-books/ When you don't have enough

control over your server to set up a robots.txt file, you can try adding a META tag to the head section of any HTML document. In example, a tag like the following tells robots not to index and not to follow links on a particular page: meta name="ROBOTS" content="NOINDEX, NOFOLLOW" Support for the META tag among robots is not so frequent as the Robots Exclusion Protocol, but most of major web indexes currently support it. NEWS POSTINGS If you want to keep the search engines out of your news postings, you can create an "X-no-archive" line in of your postings' headers:



X-no-archive: yes But although common news clients, allow you to add an X-no-archive line to the headers of your news postings, some of them dont permit you to do so. The problem is that most search engines assume that all information they find is public unless marked otherwise. So be careful because though the robot and archive exclusion standards may help keep your material out of major search engines there are some others **seo services** that respect no such rules. If you're highly concerned about the privacy of your e-mail and Usenet postings, you must use some anonymous remailers and PGP. You can read about it here:
<http://www.well.com/user/abacard/remail.html> <http://www.io.com/~combs/htmls/crypto.html>
<http://world.std.com/~frnl/pgp/> Even if you are not particularly concerned about privacy, remember that anything you write will be indexed and archived somewhere for eternity, so use the robots.txt file as much as you need it.