

چیست؟ Robots.txt فایل

همانطور که همه ما می دانیم، ربات های موتورهای جستجو برجسته ترین نقش را در آشنایی کامل یک سایت برای موتورهای جستجو دارند. گوگل به عنوان بزرگترین مرجع و گسترده ترین شبکه در این زمینه، بسیاری از ربات های خصوصی و عمومی را توسعه داده است. وب مسترهای موفق همیشه عملکرد و تغییرات این ربات ها را دنبال می کنند و طبق استانداردهای آنها حرکت می کنند.

اما چگونه این ربات ها به سایت های مختلف دسترسی پیدا می کنند؟ چگونه می توانیم دسترسی این ربات ها را به مطالب یا صفحات خاصی از سایت خود محدود است که می Robots.txt کنیم یا به آنها دسترسی کامل بدهیم؟ برای این، یک تکنیک ساده اما فوق العاده مهم وجود دارد. این تکنیک استفاده از یک فایل متنی به نام Robots.txt است که می تواند برای درج دستورات در وب برای انجام کارهای مختلف از جمله مسدود کردن یا محدود کردن دسترسی به ربات های موتور جستجو استفاده شود.

اهمیت و بهینه سازی این فایل می تواند نقش بسیار مفیدی در پیشرفت سایت اینترنتی شما داشته باشد و برعکس بی دقتی در کار با آن به راحتی می تواند رتبه سایت آشنا کنیم و از اهمیت بسیار زیاد آنها برای موفقیت بهتر در فرآیندهای سئو Robots.txt شما را چندین برابر کند. در این مقاله قصد داریم شما را با ماهیت فایل های به شما بگوئیم.

چیست؟ Robots.txt

در واقع یک فایل متنی ساده است که در مسیر اصلی سایت یا دایرکتوری ریشه شما قرار می گیرد. در واقع وظیفه این فایل ارائه بخش های قابل Robots.txt دسترسی برای دسترسی ربات و بخش های محدود شده یا با تعریف دقیق تر، بررسی و ثبت اطلاعات برای خزنده های وب (خزنده های وب) ایجاد شده توسط موتورهای جستجو می باشد. محل کار

با قرار دادن دستورات خاص در این فایل، می توانید به ربات های موتورهای جستجو بگویید که کدام صفحات، کدام فایل ها و کدام بخش های سایت شما را جستجو و Robots.txt فهرست بندی کنند و کدام صفحات را نادیده بگیرند. به عبارت دیگر، اولین چیزی که ربات های موتورهای جستجو با آن مواجه می شوند، این فایل است. خزنده های وب به محض مواجه شدن با این فایل شروع به بررسی محتویات و لیست داخل این فایل می کنند تا قسمت های قابل دسترسی را پیدا کنند.

:همانطور که گفته شد این فایل باید در روت روت هاست شما قرار گیرد. در این صورت آدرس دسترسی به این فایل به صورت زیر خواهد بود:

<https://site.com/robots.txt>



فایل Robots.txt

```
User-agent: *
Disallow: /*?
Disallow: /*s=*
Disallow: /wp-admin/
Disallow: /author/
Disallow: /herowpheader-item/
Disallow: /herowpfooter-item/
Disallow: /search/
Allow: /wp-admin/admin-ajax.php
```

Sitemap: https://seolearn.ir/sitemap_index.xml

سوالات متداول

بدرستی آپلود نکنیم چه میشود؟ robot.txt اگر فایل

نتیجه این مورد به نوع مشکل بستگی دارد. در صورتی که فایل مذکور با فرمت استاندارد و صحیح ساخته نشده باشد و یا اطلاعات و دستورات داخل آن قابل شناسایی نباشد، ربات های موتور جستجو همچنان به اطلاعات سایت شما دسترسی پیدا کرده و ایندکس می کنند. به عبارت دیگر، ربات ها تنها زمانی رفتار خود را تغییر می دهند که دستور دقیق مربوط به آن رفتار را از طریق متن داخل این فایل دریافت کنند. در غیر این صورت به رفتار عادی خود یعنی خزیدن و ایندکس کردن تمامی قسمت های سایت ادامه می دهند.

نداشته باشیم چه میشود؟ robot.txt اگر فایل

اگر این فایل در هاست سایت شما آپلود نشود، خزنده ها و ربات های موتور جستجو به تمام صفحات عمومی دسترسی خواهند داشت و می توانند تمام محتوای سایت شما را ایندکس کنند.

یک فایل متنی با ساختار ساده است. نحوه کار این فایل با کمک دستورات پیش فرض و یکپارچه سازی کلمات کلیدی تعیین می شود. مهم ترین و robots.txt فایل هستند که در زیر هر کدام را به صورت جداگانه شرح sitemap و crawl-delay، allow، Disallow، user-agent رایج ترین دستورات مواردی مانند خواهیم داد.

این دستور برای تعیین اینکه آیا ربات ها و خریداران می توانند به قسمت های خاصی از سایت دسترسی داشته باشند یا خیر استفاده می شود. با این **User-Agent**: دستور العمل، می توانید به همه ربات ها دسترسی بدهید یا می توانید با افزودن یک نام ربات خاص، دسترسی فردی را فقط به آن ربات بدهید یا محدود کنید.

مثال: یک ربات وارد سایت شما می شود و می خواهد اطلاعات یک صفحه خاص را بررسی و جمع آوری کند، به عنوان مثال

را بررسی می کند. به عنوان مثال، محتویات این فایل به شرح robots.txt قبل از خزیدن در این صفحه، ربات ابتدا فایل www.example.com/test.html.

زیر است:

User-agent: *

به این معنی است که تمام قسمت های این سایت برای همه ربات ها و خزنده های موتورهای جستجو قابل دسترسی است. اما اگر می خواهید **User-agent: *** اطلاعات خود را فقط برای ربات خاصی مشخص کنید، باید به جای ستاره، نام دقیق آن ربات را قرار دهید.

مثال زیر را در نظر بگیرید. در این مثال، تنها روبات های گوگل حق دسترسی به صفحات سایت را دارند:

User-agent: Googlebot

یا ربات های تعیین شده بررسی و **User-agent** با کمک این دستور العمل می توان مشخص کرد که کدام قسمت های سایت توسط **Disallow and Allow**: ایندکس شوند یا نه. همانطور که مشخص است، کد اجازه برای ایجاد دسترسی و از کد انکار برای محدود کردن دسترسی ربات استفاده می شود.

قرار دهید، به روبات ها می گوئید که نباید هیچ صفحه ای را در این سایت بازدید، خزیدن، یا فهرست robots.txt را در فایل **Disallow**: مثال: اگر فقط دستور استفاده کنید **Allow**: بندی کنند. همچنین، اگر می خواهید به همه آنها دسترسی کامل بدهید، باید از دستور

می توانید فایل ها، مسیرها، آدرس ها و صفحات سایت خود را به عنوان انتخابی تنظیم کنید تا فقط این قسمت ها قابل دسترسی باشند یا برعکس. مثال ساده زیر را در نظر بگیرید:

Disallow: /wp-admin/

Allow: /contact-us/

مانند مثال بالا، مسیر دسترسی به پنل مدیریت وردپرس برای دسترسی ربات ها محدود است، اما صفحه تماس قابل دسترسی است. بنابراین، این نوع دسترسی را می توان بسیار آسان و سریع مدیریت کرد.

می توانید ربات های موتور جستجو را مجبور کنید تا با وارد کردن دستور العمل تأخیر خزیدن یا نرخ تأخیر، مدت زمانی را که برای بررسی و **Crawl-Delay**: قابل استفاده **Googlebot** فهرست بندی صفحات خود تعیین می کنید، منتظر بمانند. البته لازم به ذکر است که این دستور العمل بر روی برخی ربات ها از جمله نیست. برای فعال کردن این ویژگی در ربات گوگل، می توانید نرخ زمان این مورد را از کنسول جستجوی گوگل در قسمت تنظیمات سایت تغییر دهید.

هدف آن کاهش تعداد درخواست های مکرر ربات ها به سرور سایت شما است. چنین اقدامی کمک زیادی به تسهیل فرآیندهای نمایه سازی می کند. برای استفاده از robots.txt این دستور می توانید متن زیر را در فایل متنی قرار دهید:

Crawl-Delay: 10

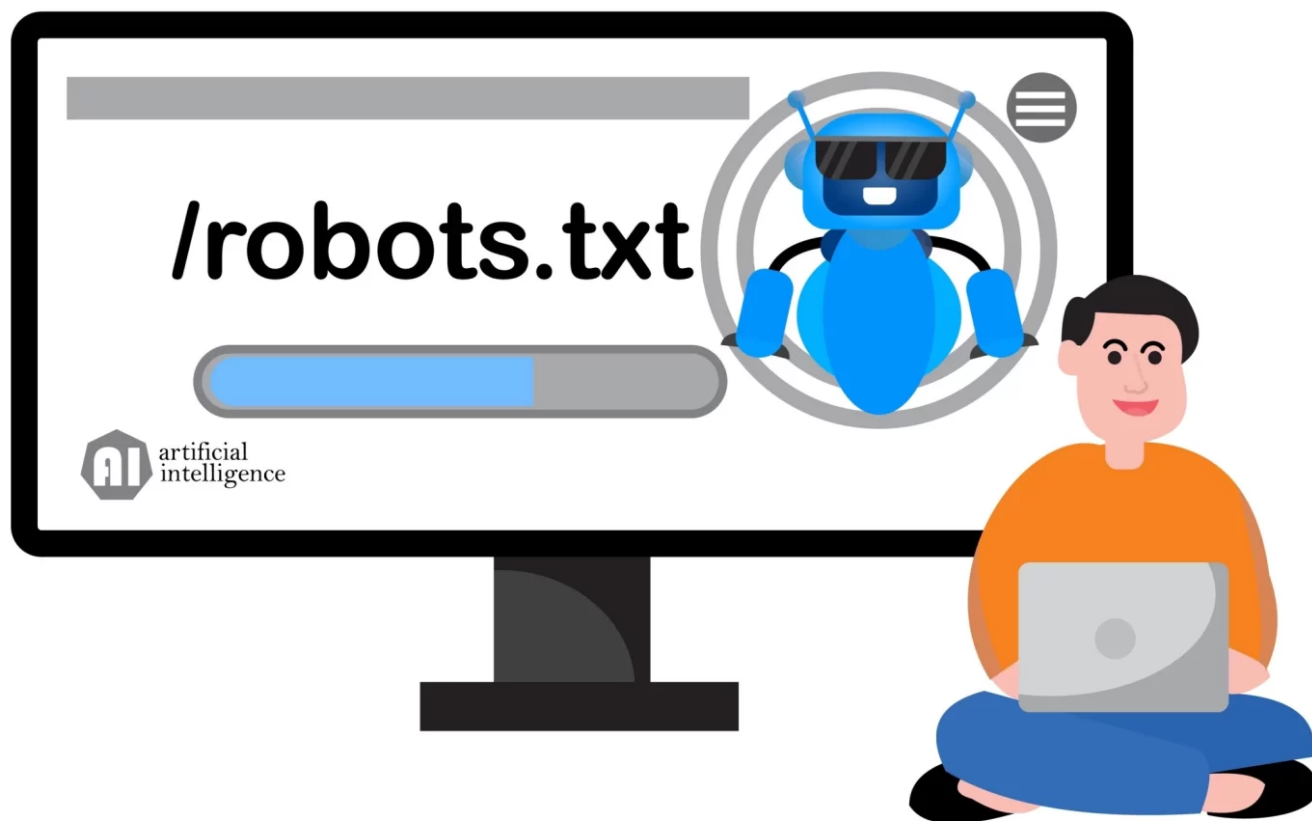
در واقع به ربات های موتورهای جستجو مانند یاهو و بینگ می گوئید که پس از هر فهرست دهی 10 ثانیه منتظر بمانند و robots.txt با قرار دادن این کد در موتور، robots.txt کمی متفاوت عمل می کند. با وارد کردن این دستور در Yandex صفحه دیگری را دوباره فهرست کنند. این کد برای موتور جستجوی هر 10 ثانیه به تمام صفحات سایت دسترسی پیدا می کند. با تغییر عدد 10 به اعداد دیگر می توانید دوره مربوطه را کم یا زیاد کنید Yandex جستجوی

ممکن است در همه موارد نتیجه درستی نداشته باشد. به عنوان مثال، اگر نرخ تاخیر ربات را crawl-delay توجه: توجه داشته باشید که استفاده از دستور العمل روی ده ثانیه تنظیم کنید، موتورهای جستجو فقط می توانند تقریباً 8600 صفحه از سایت شما را در روز بخزند و ایندکس کنند. در حالی که این مبلغ برای چنین سایت های بزرگی مقدار کمی نیست، اما برای سایت هایی با صفحات زیاد و محتوای زیاد روش مناسبی نیست. بنابراین قبل از استفاده از این محصول مراقب آن باشید.

نقشه سایت خود XML می توانید به موتورهای جستجو بگوئید که چگونه فایل robots.txt با تعیین دستورات مربوط به نقشه سایت در: (sitemap) نقشه سایت را پیدا کرده و به آن دسترسی داشته باشند. اگرچه ابزارهای ویژه ای برای ثبت نقشه سایت در موتورهای جستجو وجود دارد، اما این روش نیز یکی از روش های ساده و کم دردسر برای این مشکل است. البته در هر شرایطی استفاده از ابزارهای وبمستر موتورهای جستجو مانند سرچ کنسول گوگل در اولویت قرار خواهد داد: اضافه کنید robots.txt گرفت. اما اگر تصمیم دارید راه آسان را انتخاب کنید، فقط کد دستوری زیر را به فایل

همانطور که می بینید، باید آدرس دقیق نقشه سایت خود را مانند مثال بالا در این خط وارد کنید.

را در متن وارد کنید، محدودیت تنها Disallow: /File.php به حروف بزرگ و کوچک حساس است. به عنوان مثال، اگر کد robots.txt نکته مهم: محتوای فایل ذخیره شود، هیچ محدودیتی برای آن file.php در آن مسیر وجود داشته باشد. اما اگر فایل مورد نظر شما با نام File.php زمانی اعمال می شود که فایلی به نام به حروف کوچک و بزرگ حساس است. این فایل باید با حروف کوچک و دقیقاً به robots.txt اعمال نخواهد شد. همچنین لازم به ذکر است که نام خود فایل ذخیره شود robots.txt صورت



بسازیم؟ robots.txt چطور فایل

قبل از ایجاد این فایل، ابتدا مطمئن شوید که چنین فایل‌ای از قبل در سرور سایت شما ایجاد نشده است. برای اطلاع از این موضوع کافی است آدرس سایت خود را در فشار دهید **Enter** را تایپ کرده و `/robots.txt` وارد کرده و در انتها عبارت

مانند: <https://seolearn.ir/robots.txt>

User-Agent اگر هنگام باز کردن این صفحه با خطای 404 مواجه شدید، به این معنی است که چنین فایل‌ای ایجاد نشده است. اما اگر پس از ورود به این صفحه کد یا کنترل پنل هاست سایت خود به قسمت مدیریت فایل بروید **FTP** را مشاهده کردید به این معنی است که سایت شما این فایل را دارد. در این صورت باید از طریق روت اصلی هاست پیدا کنید. سپس باید آن را از طریق یک ویرایشگر متن آنلاین اجرا کنید و آن را ویرایش کنید. همچنین می‌توانید این `robots.txt` و فایل فایل را دانلود کرده و در کامپیوتر خود ویرایش کنید و در نهایت فایل ذخیره شده را با فایل قبلی جایگزین کنید.

با حروف `robots` ویندوز ایجاد کنید و آن را با نام **Notepad** از طریق `txt` همچنین اگر این فایل در هاست سایت شما یافت نشد، کافی است یک فایل متنی ذخیره کنید. سپس دستورات مورد نیاز خود را وارد کرده و فایل را ذخیره کنید. در نهایت فایل آماده شده را با همان روش هایی که در بالا توضیح داده شد (کوچک در روت اصلی سایت آپلود کنید (یا فایل منیجر کنترل پنل هاست **FTP** از طریق)

مقالات پیشنهادی:

سنو کلاه سفید چیست؟

افزونه یا پلاگین چیست؟

الگوریتم موبایل گدن چیست؟

شما چه امتیازی به این مقاله می‌دهید؟