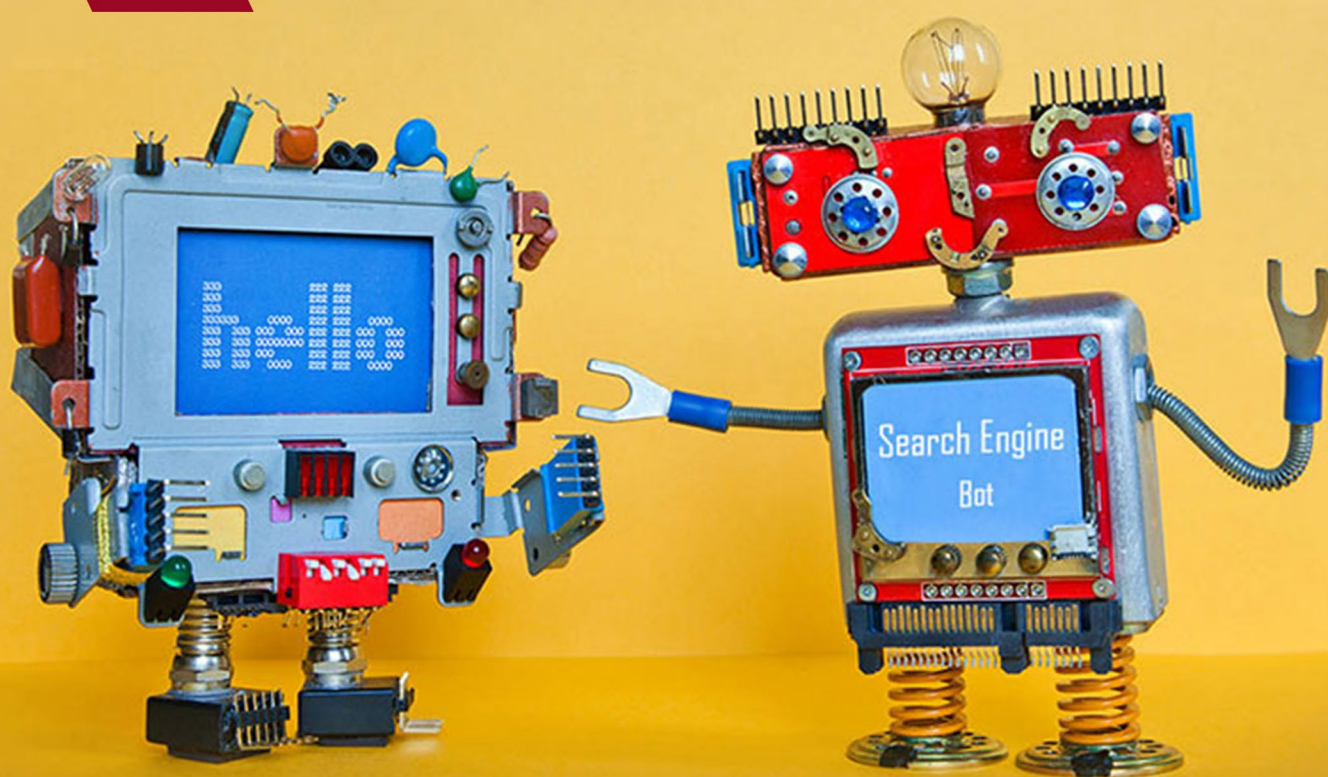


فایل Robots.txt چیست

و چه کاربردی در سئوی سایت دارد؟



نویسنده: مهران منصوری فر

در مورد Robots.txt چه می‌دانید؟ آیا می‌دانید که فایل robots.txt در چه مواردی کاربرد دارد؟ با مدیروب همراه باشید تا در این مقاله به معرفی و بررسی این مقوله مهم که فایل Robots.txt چیست و چه نقشی در بحث بهینه‌سازی سایت‌ها در موتورهای جستجو دارد می‌پردازیم.

همانطور که همه‌ی ما می‌دانیم، ربات‌های موتورهای جستجو پررنگ‌ترین نقش ممکن برای معرفی کامل یک سایت به موتورهای جستجو را ایفا می‌کنند. گوگل به عنوان بزرگترین مرجع و گسترده‌ترین شبکه در این زمینه، ربات‌های اختصاصی و عمومی زیادی گسترش داده است. وب‌مسترهای موفق همواره عملکرد و تغییرات این ربات‌ها را دنبال کرده و مطابق با استانداردهای آنان پیشروی می‌کنند.

اما این ربات‌ها به چه صورت به سایت‌های گوناگون دسترسی پیدا می‌کنند؟ چطور می‌توان دسترسی این ربات‌ها به محتوا یا صفحات خاصی از سایتمان را محدود کنیم یا به آن‌ها دسترسی کامل بدهیم؟ برای این مورد، تکنیکی ساده و در عین حال فوق‌العاده مهم و حیاتی وجود دارد. این تکنیک، استفاده از یک فایل متنی موسوم به Robots.txt است که با کمک آن و درج دستورات تحت وب می‌توان کارهای مختلفی از جمله ایجاد دسترسی یا محدودسازی ربات موتورهای جستجو را انجام داد.

اهمیت و بهینه‌سازی این فایل می‌تواند نقش بسیار مفیدی در پیشرفت سایت اینترنتی شما داشته باشد و درست برعکس، بی‌توجهی و بی‌دقتی در کار با آن، ممکن است به راحتی رتبه سایت شما را چندین برابر بدتر کند. ما در این مقاله قصد داریم تا شما را با ماهیت فایل‌های Robots.txt آشنا کرده و اهمیت بسیار زیاد آن‌ها برای کسب موفقیت بهتر در فرآیندهای سئو به شما بازگو کنیم.

[بیشتر بدانید: موتورهای جستجو چگونه کار می‌کنند؟](#)

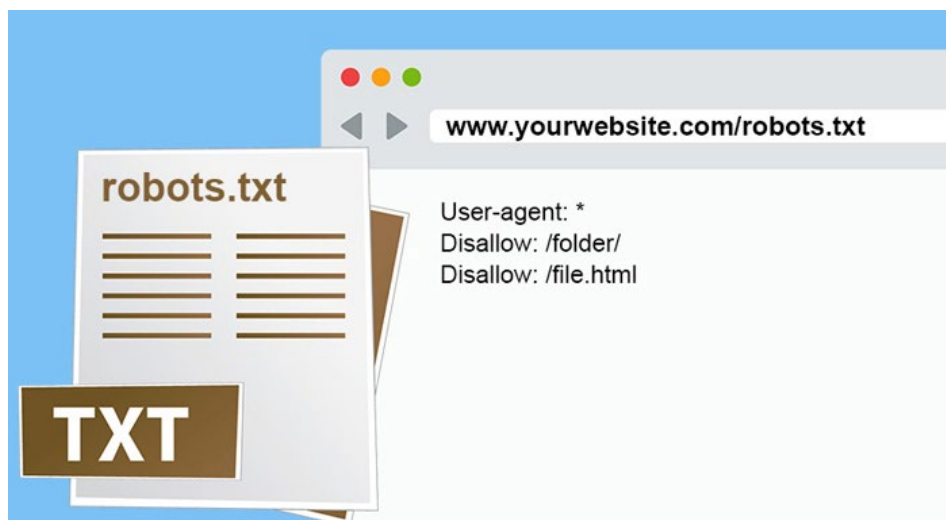
فایل Robots.txt چیست؟

یک Robots.txt در واقع یک فایل متنی ساده است که در مسیر اصلی فضای سایت شما یا روت دایرکتوری (Root Directory) قرار می‌گیرد. در حقیقت وظیفه این فایل معرفی بخش‌های قابل دسترسی و بخش‌های محدود شده برای دسترسی ربات‌ها یا با تعریفی دقیق‌تر، خزندگان سطح وب (Web Crawlers) که از سوی موتورهای جستجو به منظور بررسی و ثبت اطلاعات سایت‌ها ایجاد شده اند می‌باشد.

با درج دستوراتی خاص در این فایل، شما می‌توانید به ربات‌های موتورهای جستجو بگوئید که کدام صفحات، کدام فایل‌ها و کدام بخش‌های سایت شما را دیده و آن‌ها را ایندکس کنند و کدام صفحات را نادیده بگیرند. به عبارت دیگر، اولین چیزی که ربات‌های موتورهای جستجو با آن برخورد می‌کنند، همین فایل Robots.txt است. به محض رویارویی با این فایل، خزندگان سطح وب شروع به بررسی محتوا و لیست داخل این فایل کرده تا بخش‌های قابل دسترسی را پیدا کنند.

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

www.YourWebsite.com/robots.txt



اگر سایت شما فایل robots.txt را نداشته باشد چه اتفاقی می‌افتد؟

اگر این فایل در هاست سایت شما آپلود نشده باشد، خزندگان و ربات‌های موتورهای جستجو امکان دسترسی به تمام صفحات عمومی را خواهند داشت و قادر هستند تا تمام محتوای سایت شما را ایندکس کنند.

اگر فایل robots.txt به درستی تهیه و آپلود نشود چه می‌شود؟

نتیجه این مورد بستگی به نوع مشکل خواهد داشت. اگر فایل مذکور با فرمت استاندارد و صحیح ایجاد نشده باشد و یا اطلاعات و دستورات داخل آن قادر به شناسایی نباشند، ربات‌های موتورهای جستجو همچنان به دسترسی به اطلاعات سایت شما ادامه داده و می‌توانند آن‌ها را ایندکس کنند. به عبارت دیگر، ربات‌ها فقط زمانی تغییر رفتار می‌دهند که دستور دقیق و منطبق با آن رفتار را از طریق متون داخل این فایل دریافت کرده باشند. در غیر این صورت آن‌ها به رفتار طبیعی خود یعنی بررسی و ایندکس کردن تمام بخش‌های سایت ادامه خواهند داد.

برخی از مهم‌ترین دلایل استفاده از robots.txt

۱. با کمک این فایل می‌توان دسترسی موتورهای جستجو را مدیریت کرد

محدود کردن دسترسی خزندگان موتورهای جستجو به صفحات، فایل‌ها و محتوای انتخابی سایت‌ها رایج‌ترین دلیل استفاده از فایل robots.txt است.

اگر برای شما این سؤال پیش آمده است که چرا باید بعضی از صفحات یا فایل‌ها را ایندکس نکنیم، در جواب باید گفت که در بسیاری از موارد، ایندکس شدن و معرفی شدن یک صفحه یا فایل از یک سایت در موتورهای جستجو می‌تواند نتایج نامناسبی به همراه داشته باشد.

به عنوان مثال شاید یک وب‌مستر قصد انتشار مطلبی خاص را داشته باشد

که مخاطبین آن مطلب نیز افرادی خاص و منتخب هستند. این مطلب ممکن است از لحاظ بعضی شرایطی، برخی از قوانین موتورهای جستجو را نقض کند و یا شامل محتوایی باشد که امکان معرفی آن به صورت عمومی وجود نداشته باشد. در این صورت می‌توان با کمک فایل Robots.txt دسترسی موتورهای جستجو به این صفحات یا فایل‌ها را محدود کرد.

از بارزترین مثال‌ها در این زمینه، سایت‌های اشتراک فایل غیرقانونی از جمله سایت‌های تورنت هستند. خیلی از این سایت‌ها اجازه دسترسی به محتوای داخلی خودشان را به ربات‌های موتورهای جستجو نمی‌دهند و به جای آن جستجوگر داخلی خودشان را به کاربران پیشنهاد می‌کنند. چون اگر محتوای این سایت‌ها توسط ربات‌ها شناسایی شوند، این سایت‌ها به خاطر ارائه محتوای غیرقانونی نه تنها رتبه و جایگاه خود را در موتورهای جستجو از دست می‌دهند، بلکه مسائل مربوط به حقوق کپی‌رایت و مسائلی از این قبیل نیز گریبان‌گیرشان خواهد شد.

۲. عدم استفاده از فایل robots.txt می‌تواند به افت عملکرد بهینه سایت دامن بزند

وبسایت‌های فعال به ویژه وبسایت‌های پر بازدید و بزرگ، روزانه هزاران بار توسط ربات‌های گوناگون از سمت موتورهای جستجو مورد بازدید و بررسی قرار می‌گیرند. هر ربات یا به اصطلاح خزنده (Crawler) طی یک فرآیند دومارحله‌ای (بررسی و سپس ایندکس) اقدام به جمع‌آوری اطلاعات از سایت‌ها می‌کند. این فرآیند شامل بررسی تمام بخش‌های سایت شما است. حالا فرض کنید به صورت روزانه صدها یا هزاران ربات شروع به بررسی و جمع‌آوری اطلاعات از سایت شما می‌کنند و در این صورت، عملکرد بهینه سایت شما و سرعت بارگذاری اطلاعات آن برای مرورگرها تحت‌الشعاع قرار خواهد گرفت.

بدیهی‌ست که رفت و آمد این حجم از ربات‌ها می‌تواند تا حد قابل توجهی

بازدهی کلی سایت را با شرایط نامناسبی همراه کند. البته اگرچه این موضوع برای سایت‌هایی که بازدید کمتری دارند بسیار کمتر به چشم می‌آید اما برای سایت‌های پر بازدید که به خودی خود باید روزانه ترافیک بازدید از سمت هزاران کاربر را هندل کنند، اضافه شدن ترافیک زیاد این ربات‌ها نیز ممکن است در دسرساز باشد.

در این مواقع اغلب وب‌مسترها به راحتی و با استفاده از robots.txt دسترسی ربات‌های موتورهای جستجو را به بخش‌های مختلف و مشخص شده‌ای که اهمیت چندانی برای سئو و کسب رتبه در موتورهای جستجو ندارند را محدود می‌کنند. در این حالت نه تنها سرور سایت با ترافیک سبک‌تری فعالیت خواهد کرد، بلکه مراحل بررسی و جمع‌آوری اطلاعات و سپس ایندکس کردن آن‌ها توسط ربات‌ها نیز به مقدار بسیار زیادی سریع‌تر خواهد شد.

۳. استفاده از فایل robots.txt می‌تواند در مدیریت لینک‌ها کاربردی باشد

از دیگر مزایای استفاده از robots.txt امکان مدیریت لینک‌ها و آدرس صفحات (URL) است. در بحث سئو مسئله‌ای با عنوان مخفی‌سازی آدرس صفحات یا URL Cloacking وجود دارد. این بحث در واقع نوعی تکنیک سئو برای پنهان کردن آدرس صفحات از دید کاربران و یا موتورهای جستجو است. با کمک robots.txt می‌توان این مدل لینک‌ها را مدیریت کرده و آدرس آن‌ها را پنهان کرد.

بیشترین کاربرد این مورد در استفاده از لینک‌های مربوط به بحث «سیستم همکاری در فروش» یا «Affiliate Marketing» می‌باشد. در این حالت، شما می‌توانید لینک‌های ایجاد شده در سیستم Affiliate که با عنوان Affiliate Links شناخته می‌شوند را مدیریت کرده و آدرس آن‌ها را مخفی کنید تا کاربران به نوعی مجبور به کلیک روی آن‌ها شوند.

توجه کنید که این تکنیک فقط باید توسط افراد حرفه‌ای انجام شود. چراکه URL Cloacking یکی از تکنیک‌های سئوی کلاه سیاه است و در صورتی که به درستی آن را پیاده‌سازی نکنید، متهم به نقض قوانین موتورهای جستجو شده و شامل جریمه‌های گوگل خواهید شد.

[بیشتر بدانید: آشنایی با خطاهای گوگل وبمستر تولز و راهنمای رفع آنها](#)



فایل robots.txt چگونه کار می‌کند؟

فایل robots.txt یک فایل متنی با ساختاری ساده است. نحوه عملکرد این فایل با کمک دستورات پیش‌فرض و ادغام کلمات کلیدی از پیش تعیین شده است. از جمله مهم‌ترین و رایج‌ترین این دستورات مواردی مثل User-agent، Crawl-delay، Allow، Disallow، Sitemap می‌باشند که در ادامه به صورت جداگانه هر یک از این موارد را با جزئیات آن‌ها شرح می‌دهیم.

User-agent: این دستور برای مشخص کردن ربات‌ها و خردنگانی است

که امکان دسترسی به بخش‌های سایت را داشته باشند یا خیر. با این دستورالعمل می‌توان به تمام ربات‌ها دسترسی داد و یا با اضافه کردن نام رباتی خاص، فقط به آن ربات دسترسی‌های مختلف را داد یا محدود کرد.

مثال: یک ربات وارد سایت شما شده و قصد بررسی و جمع‌آوری اطلاعات از یک صفحه خاص برای مثال `www.example.com/test.html` را دارد. این ربات قبل از اینکه این صفحه را مورد بررسی قرار دهد، ابتدا فایل `robots.txt` را چک می‌کند. برای مثال محتویات داخل این فایل به صورت زیر است:

*:User-agent

:User-agent * به این معنی است که تمام بخش‌های این سایت قابل دسترس برای تمام ربات‌ها و خزندگان موتورهای جستجو است. اما اگر قصد این را دارید که اطلاعات خود را فقط برای یک ربات خاص تعیین کنید باید نام دقیق آن ربات را به جای ستاره جایگذاری کنید.

به مثال زیر دقت کنید. در این مثال فقط ربات گوگل حق دسترسی به صفحات سایت را دارد:

User-agent: Googlebot

Disallow و Allow: با کمک این دستورالعمل می‌توان به User-agent یا همان ربات‌های تعیین شده مشخص کرد که کدام بخش‌های سایت را بررسی و ایندکس کنند یا نکنند. همانطور که مشخص است، کد Allow به منزله ایجاد دسترسی و کد Disallow به منزله محدودسازی دسترسی ربات‌ها استفاده می‌شوند.

مثال: اگر فقط دستور «Disallow: /» را در فایل `robots.txt` درج کرده باشید،

با این کار به ربات‌ها اعلام می‌کنید که آن‌ها نباید هیچ صفحه‌ای از این سایت را بازدید، بررسی یا ایندکس کنند. همچنین اگر می‌خواهید به همه آن‌ها دسترسی کامل داده باشید باید از دستور «Allow: /» استفاده کنید.

شما می‌توانید فایل‌ها، مسیرها، آدرس‌ها و صفحات مشخص شده سایت خود را به عنوان موارد انتخابی تعیین کنید تا فقط این بخش‌ها قابل دسترسی باشند یا برعکس. به نمونه ساده زیر دقت کنید:

/Disallow: /wp-admin

/Allow: /contact

طبق مثال بالا، مسیر ورودی به پنل مدیریتی وردپرس برای دسترسی ربات‌ها محدود شده است اما صفحه Contact یا همان صفحه ارتباط با ما قابل دسترس است. به همین طریق می‌توان خیلی راحت و سریع این نوع دسترسی‌ها را مدیریت کرد.

نمونه‌هایی رایج از این دستورالعمل‌ها :

فایل Robots.txt چیست و چه کاربردی در سئوی سایت دارد؟

دستورالعملها (قوانین)	مثال (نمونه)
<p>عدم دسترسی به تمامی محتوای سایت</p> <p>با این کد دسترسی تمامی رباتها به بررسی تمامی بخشهای سایت از بین می‌رود. به یاد داشته باشید که در بعضی از شرایط، برخی از آدرس صفحات سایت شما (URL) حتی با استفاده از این دستورالعمل نیز باز هم قابل ایندکس شدن هستند. بنابراین اگر مایلید صفحه‌ای از سایت خود به هیچ وجه در موتورهای جستجو قرار نگیرد باید آن صفحه را به صورت noindex و در موارد خاص، برای آن صفحه رمز عبور قرار دهید. در صورت ثبت رمز عبور، هیچ رباتی قادر به بررسی محتوای داخل آن صفحه نخواهد بود.</p>	<p>User-agent: *</p> <p>Disallow: /</p>
<p>محدود کردن دسترسی به مسیریهای خاص</p> <p>استفاده از این دستورات به منظور محدود کردن دسترسی رباتها به مسیریهای خاص است. کافیسیت بعد از: Disallow: یک / گذاشته و نام مسیر مورد نظر خود را تایپ کنید و آن را با یک / دیگر ببندید.</p> <p>توجه کنید که درست مثل مثال ذکر شده، برای محدود کردن هر صفحه باید یک خط جدید بسازید و برای هرکدام یک دستور مشابه ایجاد کنید.</p>	<p>User-agent: *</p> <p>Disallow: /calendar/</p> <p>Disallow: /junk/</p>
<p>ایجاد دسترسی برای یک ربات (خرزنده) خاص</p> <p>با این دستور می‌توان فقط یک ربات منتخب را استثنا قرار داد تا بتواند به محتوای صفحات دسترسی داشته باشد.</p>	<p>User-agent: Googlebot-news</p> <p>Allow: /</p> <p>User-agent: *</p> <p>Disallow: /</p>
<p>ایجاد دسترسی برای همه رباتها (خرزندگان) به جز یک ربات</p> <p>برعکس مورد بالا، امکان ایجاد دسترسی به تمامی رباتها به جز رباتهای انتخابی شما نیز وجود دارد که با استفاده از این دستورالعمل می‌توان این دسترسیها را ایجاد کرد.</p>	<p>User-agent: Unnecessarybot</p> <p>Disallow: /</p> <p>User-agent: *</p> <p>Allow: /</p>
<p>محدود کردن دسترسی به صفحات خاص</p> <p>با این دستورالعمل می‌توانید دسترسی رباتها به یک صفحه وب را محدود کنید. برای محدود کردن صفحات بیشتر، باید همین کد را مجدداً نوشته و همانند نمونه، بعد از علامت / آدرس صفحه مورد نظر با پسوند مربوطه را وارد کنید.</p>	<p>Disallow: /private_file.html</p>
<p>مسدود کردن یک تصویر خاص در موتور جستجوی گوگل</p> <p>برای مسدودسازی آدرس یک تصویر (آدرس آپلود شده عکس) می‌توان از این کد دسترسی استفاده کرد تا این عکس در بخش نتایج جستجوهای تصاویر گوگل نمایش داده نشود. همانطور که مشخص است، باید آدرس مسیر تصویر را به صورت نمونه وارد کنید.</p>	<p>User-agent: Googlebot-Image</p> <p>Disallow: /images/dogs.jpg</p>
<p>مسدودسازی تمامی تصاویر روی سایت در موتور جستجوی گوگل</p> <p>استفاده از این دستور منجر به مسدود شدن دسترسی ربات گوگل به ایندکس کردن تمامی تصاویر موجود در سایت شما خواهد شد.</p>	<p>User-agent: Googlebot-Image</p> <p>Disallow: /</p>
<p>مسدودسازی فرمتی خاص از فایلها برای ایندکس شدن در موتور جستجوی گوگل</p> <p>با کمک این کد می‌توانید به راحتی فرمت‌های مورد نظر خود که نمی‌خواهید در نتایج صفحات موتور جستجوی گوگل به نمایش در بیایند را انتخاب کرده و دسترسی ربات گوگل به آن‌ها را مسدود کنید. کافیسیت که برای هر فرمت، یک دستورالعمل مثل دستور مشابه ایجاد کرده و فرمت مورد نظر خود در آن جایگزین کنید. برای مثال در این نمونه ما فرمت تصاویر jpg را مسدودسازی کرده‌ایم.</p>	<p>User-agent: Googlebot</p> <p>Disallow: /*.jpg\$</p>
<p>محدود کردن دسترسی رباتها به کل سایت، به جز تبلیغات AdSense گوگل</p> <p>با این دستور می‌توان دسترسی تمام خرزندگان به جز رباتهای طرف قرارداد شبکه تبلیغاتی گوگل به سایت را محدود کرد. با این کار صفحات شما از نتایج جستجوی موتورهای جستجو حذف می‌شوند اما رباتها (خرزندگان) مربوط به شبکه تبلیغاتی گوگل که با نام Mediapartners-Google شناخته می‌شوند همچنان می‌توانند تبلیغاتی که به صفحات سایت شما لینک داده شده اند را به بازدیدکنندگان نمایش دهند.</p>	<p>User-agent: *</p> <p>Disallow: /</p> <p>User-agent: Mediapartners-Google</p> <p>Allow: /</p>
<p>مسدودسازی URL هایی که به فرمت‌هایی خاص ختم می‌شوند</p> <p>درست مثل کدی که برای مسدودسازی فرمت فایلها بود، می‌توان از این کد برای محدودسازی دسترسی رباتها به پسوندهای خاص در آدرس صفحات استفاده کرد.</p>	<p>User-agent: Googlebot</p> <p>Disallow: /*.pdf\$</p>

Crawl-delay: شما قادر هستید تا با درج دستورات عمل Crawl-delay یا نرخ تأخیر، ربات‌های موتورهای جستجو را مجبور کنید تا برای بررسی و ایندکس کردن صفحات شما به نوبت، به مقدار زمان مشخص شده توسط شما صبر کنند. البته قابل ذکر است که این دستورات عمل روی بعضی از ربات‌ها از جمله مهمترین آن‌ها ربات گوگل یا Googlebot قابل استفاده نیست. برای اینکه این قابلیت را در ربات گوگل نیز فعال کنید، می‌توانید از کنسول جستجوی گوگل (Google Search Console) نرخ زمان مربوط به این مورد را برای ربات گوگل از بخش تنظیمات سایت (Site Settings) تغییر دهید.

هدف از این کار کاهش تعداد درخواست‌های پی‌پی ربات‌ها به سرور سایت شما است. چنین اقدامی به تسهیل فرآیندهای ایندکس شدن کمک زیادی می‌کند. برای استفاده از این دستور می‌توانید در فایل متنی robots.txt متن زیر را قرار دهید :

crawl-delay: ۱۰

با قرار دادن این کد در robots.txt شما در واقع به ربات‌های موتورهای جستجوی همچون Yahoo و Bing اعلام می‌کنید که بعد از هر یک ایندکس، ۱۰ ثانیه صبر کرده و مجدداً صفحه دیگری را ایندکس کنند. این کد برای موتور جستجوی Yandex با کمی تفاوت عمل می‌کند. با درج این دستور در robots.txt موتور جستجوی Yandex هر ۱۰ ثانیه یک بار به کل صفحات سایت دسترسی پیدا خواهد کرد. با تغییر عدد ۱۰ به دیگر اعداد، می‌توانید مدت زمان مربوطه را کم یا زیاد کنید.

نکته: توجه کنید که استفاده از دستورات عمل crawl-delay ممکن است در همه حالات نتیجه مناسب به همراه نداشته باشد. به عنوان مثال اگر شما نرخ تأخیر ربات‌ها را روی ده ثانیه قرار دهید، موتورهای جستجو به صورت روزانه می‌توانند فقط نزدیک به ۸۶۰۰ صفحه از سایت شما را بررسی و ایندکس کنند.

اگرچه این مقدار برای سایت‌های نه چندان بزرگ مقدار کمی نیست، اما برای سایت‌هایی که صفحات زیاد و محتوای متعددی دارند اقدام مناسبی نمی‌باشد. بنابراین قبل از استفاده از این مورد به این مسئله دقت کافی را داشته باشید.

Sitemap: با تعیین دستور مربوط به نقشه سایت در robots.txt می‌توانید به موتورهای جستجو بفهمانید که فایل XML نقشه سایت شما را از چه مسیری پیدا کرده و به آن دسترسی داشته باشند. اگرچه برای ثبت کردن نقشه سایت در موتورهای جستجو ابزارهای انحصاری خاصی وجود دارد، ولی این راه نیز یکی از روش‌های ساده و کم‌دردسر برای این مسئله است. البته تحت هر شرایطی، استفاده از ابزارهای وب‌مستری موتورهای جستجو مثل Google Webmaster Tools اولویت بسیار بیشتری خواهد داشت. اما اگر تصمیم به انتخاب راه ساده‌تر دارید، کافست کد دستوری زیر را در فایل robots.txt اضافه کنید :

Sitemap: <https://example.com/sitemap.xml>

همانطور که پیداست، باید مثل نمونه بالا، آدرس دقیق نقشه سایت خود را به صورت صحیح در این لاین وارد کنید.

نکته بسیار مهم: محتوای فایل robots.txt به بزرگی یا کوچکی حروف حساس است. برای مثال اگر شما کد Disallow: /File.php را در متن قرار دهید فقط در صورتی که فایلی با عنوان File.php در آن مسیر وجود داشته باشد محدودیت روی آن اعمال می‌شود؛ اما اگر فایل مورد نظر شما با نام file.php ذخیره شده باشد، هیچ محدودیتی روی آن قرار نخواهد گرفت. همچنین قابل ذکر است که نام خود فایل robots.txt نیز به بزرگی و کوچکی حروف حساس است. این فایل حتماً باید با حروف کوچک و دقیقاً به صورت robots.txt ذخیره شود.

چطور یک فایل robots.txt بسازیم؟

قبل از ساخت این فایل، اول مطمئن شوید که پیش از این چنین فایلی در سرور سایت شما ایجاد نشده باشد. برای اینکه از این مورد آگاه شوید، کافیست آدرس سایت خود را وارد کرده و در انتهای آن robots.txt/ را تایپ کرده و اینتر بزنید.

یعنی به این صورت: www.Modireweb.com/robots.txt

اگر در صورت باز کردن این صفحه با ارور ۴۰۴ مواجه شوید یعنی چنین فایلی ساخته نشده است. اما در صورتی که بعد از وارد شدن به این صفحه با کدهای User-agent و ... مواجه شدید، یعنی سایت شما این فایل را دارد. در این حالت باید از طریق FTP و یا کنترل پنل هاست سایت خود، به قسمت مدیریت فایل رفته و در روت اصلی هاست فایل robots.txt را پیدا کنید. سپس باید آن را توسط یک ویرایشگر متنی آنلاین اجرا کرده و آن را ویرایش کنید. همچنین می‌توانید این فایل را دانلود کرده و روی کامپیوتر خود ویرایش کنید و در نهایت فایل ذخیره شده را با فایل قبلی جایگزین کنید.

همچنین اگر این فایل در هاست سایت شما پیدا نشد، خیلی راحت یک فایل متنی txt از طریق Notepad ویندوز بسازید و آن را با نام robots (با حروف کوچک) ذخیره کنید. سپس دستورات مورد نیاز خود را درون آن قرار داده و فایل را ذخیره سازی کنید. در نهایت، فایل آماده شده را از همان روش‌های گفته شده (از طریق FTP یا فایل منیجر کنترل پنل هاست) در روت اصلی سایت آپلود کنید.

چطور فایل robots.txt خودمان را تست کرده و از عملکرد صحیح آن مطمئن شویم؟

در حالت عادی و با باز کردن آدرس فایل robots.txt و بررسی دستی دستورات

داخل آن می‌توانید از صحت صحیح بودن دستورات مطمئن شوید. اما راه حرفه‌ای تر استفاده از ابزار تست فایل robots.txt کنسول جستجوی گوگل است. راهنمایی بیشتر:

برای این منظور ابتدا وارد سایت Google Search Console شوید روی بخش robots.txt Tester کلیک کنید گزینه Test را انتخاب کنید

اگر مشکلی در فایل وجود نداشته باشد، دکمه قرمز رنگ Test به دکمه سبز رنگ Allowed تغییر پیدا می‌کند. اما اگر ایرادی در فایل دیده شود، لاین (دستور) مشکل دار هایلایت شده و مشخص می‌شود.

همچنین با کمک این ابزار گوگل می‌توان از قابلیت‌های دیگری نیز بهره‌مند شد. مثلاً می‌توانید از بخش URL Tester آدرس صفحات سایت خود را آنالیز کرده تا متوجه شوید کدام یک از آن‌ها برای ربات‌ها مسدود شده هستند.

همچنین این ابزار یک ویرایشگر آنلاین نیز می‌باشد که می‌توانید محتویات آن را با استانداردهای پیشنهادی تغییر دهید. اما فراموش نکنید که به منظور اعمال این تغییرات در سایت خود، باید محتوای ویرایش شده جدید را با محتوای متنی قبلی موجود در robots.txt حال حاضر در هاست خود جایگزین کنید.

آموزش پیدا کردن robots.txt Tester در Google Search Console

فایل Robots.txt چیست و چه کاربردی در سئوی سایت دارد؟

Dashboard
Messages
Search Appearance
Search Traffic
Google Index
Crawl
Crawl Errors
Crawl Stats
Fetch as Google
robots.txt Tester
Sitemaps
URL Parameters
Security Issues
Other Resources

robots.txt Tester

Edit your robots.txt and check for errors. [Learn more.](#)

Latest version seen on 1/16/16, 8:25 PM OK (200) 67 Bytes [See live robots.txt](#)

```
1 User-agent: *
2 Disallow: /wp-admin/
3 Allow: /wp-admin/admin-ajax.php
4
```

0 Errors 0 Warnings

Submit

http://example.com/ Enter a URL to test if it is blocked Googlebot TEST

مشکل نداشتن کدهای دستوری در robots.txt و تأیید مجوز (Allowed)

robots.txt Tester

Edit your robots.txt and check for errors. [Learn more.](#)

Latest version seen on 5/2/16, 9:18 AM OK (200) 134 Bytes [See live robots.txt](#)

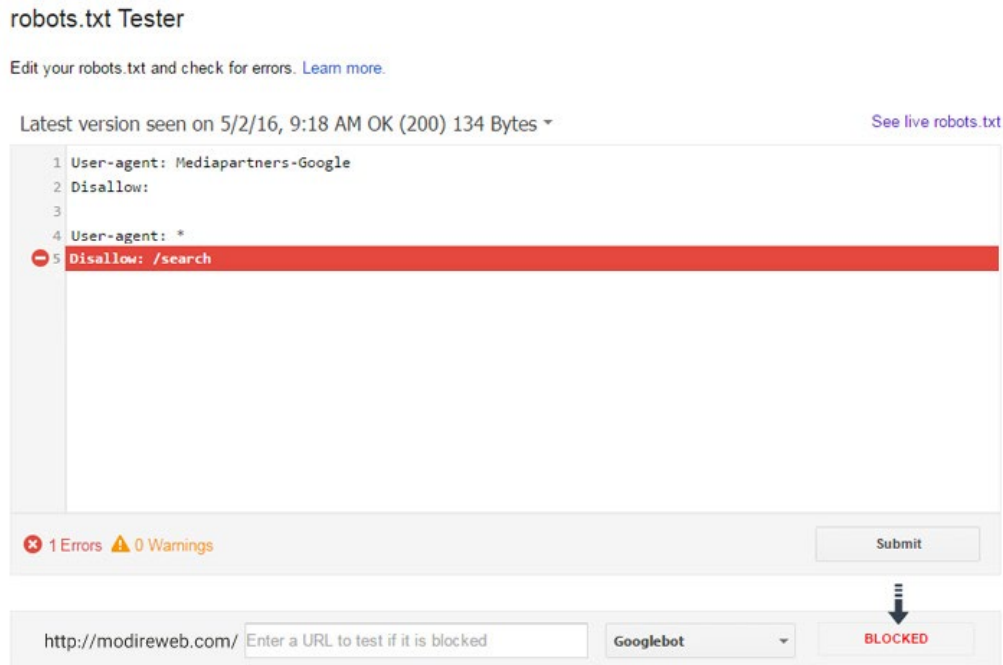
```
1 User-agent: Mediapartners-Google
2 Disallow:
3
4 User-agent: *
5 Disallow: /search
6 Allow: /
```

0 Errors 0 Warnings

Submit

http://modireweb.com/ Enter a URL to test if it is blocked Googlebot ALLOWED

مشکل داشتن کدهای دستوری در robots.txt و عدم تأیید مجوز (Disallowed)



گوگل برای این موضوع تدابیر ساده‌ای اندیشیده است. کفایت بعد از اعمال تغییرات و ویرایش متن فایل robots.txt با کمک این ابزار، فرآیند مذکور را با تأیید کردن (کلیک روی دکمه Submit) ادامه دهید تا به بخش بعدی هدایت شوید. در بخش بعد که شامل ۳ بخش است، شما می‌توانید فایل robots.txt ویرایش شده جدید را دانلود کنید. سپس باید فایل جدید را با فایل قبلی در هاست خود جایگزین کنید. بعد از این کار مجدداً به این صفحه برگشته و گزینه Submit نهایی را بزنید تا به گوگل اطلاع دهید که فایل مورد نظر بروزرسانی شده است.

- 1** Download updated code
Download the updated robots.txt from the editor. [Download](#)
- 2** Check uploaded version
Upload the updated robots.txt to your domain's root, then check that your uploaded robots.txt is the version you want Google to use. [View uploaded version](#)
- 3** Ask Google to update
Submit a request to let Google know your robots.txt file has been updated. [Submit](#)

شرایط Robots.txt در وردپرس به چه صورت است؟

تمام مواردی که در بالا گفته شد، برای سیستم مدیریت محتوای Wordpress نیز صدق می‌کند. اما چند نکته راجع به وردپرس در این زمینه وجود دارد که دانستن آن‌ها خالی از لطف نیست.

نکته اول:

در گذشته پیشنهاد میشد که صفحه پنل مدیریتی وردپرس یا همان wp-admin را از طریق فایل robots.txt برای دسترسی ربات‌ها مسدودسازی کنیم. اما بعد از بروزرسانی وردپرس در سال ۲۰۱۲ این مسئله دیگر اهمیتی نداشت. چرا که وردپرس در سیستم جدید خود قابلیتی پیش‌فرض به تگ هدر فایل robots.txt اضافه کرده بود که در آن تگ noindex و دستور مسدودسازی دسترسی (disallow) گنجانده شده بود. بنابراین در حال حاضر بدون نیاز به مورد خاصی، صفحه مدیریتی پنل سایت شما به صورت خودکار از لیست ایندکس‌های موتورهای جستجو حذف می‌شود.

نکته دوم:

سیستم مدیریت محتوای Wordpress به صورت پیش فرض یک فایل robots.txt دارد که از این فایل با نام فایل مجازی robots.txt یا Virtual robots.txt یاد می شود. این به این معنی است که شما نمی توانید به صورت مستقیم این فایل را پیدا کرده و آن را ویرایش کنید. تنها راهی که بتوان این فایل را مشاهده کرد باز کردن آدرس مستقیم آن در مرورگرها (http://www.example.com/robots.txt) است.

دستورالعمل های پیش فرض موجود در فایل robots.txt مجازی در وردپرس شامل موارد زیر است:

*:User-agent

/Disallow: /wp-admin

Allow: /wp-admin/admin-ajax.php

در این حالت ربات ها به صفحه admin-ajax.php دسترسی دارند. اگر می خواهید بدون دستکاری خاصی این حالت را نیز غیرفعال کنید، باید به تنظیمات وردپرس در صفحه wp-admin رفته و گزینه Search Engine Visibility را فعال کنید تا دسترسی تمامی ربات ها به سایت شما مسدود شود.

چطور باید فایل robots.txt در وردپرس را ویرایش کرد؟

طبق چیزی که گفته شد (عدم امکان ویرایش مستقیم فایل robots.txt در وردپرس) تنها راهی که بتوان فایل robots.txt را ویرایش کرد ساختن یک فایل جدید از آن و آپلود کردن آن در روت اصلی هاست سرور سایت است. در واقع وقتی یک فایل جدید با این عنوان ساخته شود، وردپرس به صورت خودکار فایل مجازی robots.txt را غیرفعال می کند.

نکات مهم فایل robots.txt در رابطه با سئو

همانطور که استفاده از دستورات robots.txt در خیلی از مواقع پیشنهاد می‌شود و برای کمپین سئوی سایت شما تأثیر مناسبی نیز خواهد داشت، زیاده‌روی و یا اشتباه کردن در آن نیز می‌تواند نتیجه عکس داشته باشد. به همین منظور نکات زیر را به طور دقیق مطالعه کرده و با توجه به این موارد یک فایل robots.txt مناسب و منطبق با استانداردهای سئو تهیه کنید :

قبل از هرچیز فایل robots.txt خود را بررسی و تست کنید تا مطمئن شوید بخش مهمی از سایت شما از طریق آن مسدود نشده باشد.

فایل‌ها و پوشه‌های مهم CSS و JS را مسدود نکنید. فراموش نکنید که ربات‌های گوگل درست از نگاه یک کاربر یک سایت را بررسی و ایندکس می‌کنند. در نتیجه اگر برخی از فایل‌های CSS و JS و دیگر فایل‌های مشابه را بلاک کنید، ممکن است سایت شما به صورت مناسب بارگذاری نشود و این مسئله نیز می‌تواند باعث ایجاد خطا شده و در ایندکس شدن دیگر صفحات سایت شما تأثیر منفی بگذارد.

اگر از WordPress استفاده می‌کنید، نیازی به مسدودسازی مسیرهایی همچون wp-admin و پوشه‌های wp-include نیست. وردپرس به صورت اتوماتیک این کار را از طریق تگ‌های متا انجام می‌دهد.

سعی کنید دستورات خاص زیادی برای ربات‌های خاص تعیین نکنید. پیشنهاد می‌شود اگر دستور خاصی دارید آن را با کمک کد User-agent: * روی همگی ربات‌ها اعمال کنید.

اگر مایلید که بعضی از صفحات سایت شما در موتورهای جستجو ایندکس نشوند، پیشنهاد می‌کنیم از متاتگ‌های هدر هر صفحه استفاده کنید. البته در حالی که تعداد این صفحات کم باشد یا این فرآیند برای شما دشوار نباشد.

[بیشتر بخوانید: گوگل آنالیتیکس و هر آنچه برای کار با آن باید بدانید](#)

نتیجه‌گیری

بحث robots.txt چیزی نیست که بخواهید وقت زیادی روی آن بگذارید و یا به صورت مستمر آن را مورد آزمون و خطا قرار دهید. اگر نیاز به بهره‌مند شدن از این قابلیت دارید، بهترین کار استفاده از ابزار آنلاین Google Search Console است. به کمک این ابزار خیلی راحت‌تر می‌توانید فایل robots.txt خود را مدیریت، ویرایش، اشکال‌زدایی و بروزرسانی کنید.

همچنین پیشنهاد می‌شود تا حد امکان از بروزرسانی پشت هم این فایل خودداری کنید. بهترین کار این است که درست بعد از ساخت وبسایت، یک فایل کامل و نهایی از robots.txt تهیه کنید. بروزرسانی‌های پیاپی این فایل اگرچه در روند فعالیت سایت شما به ظاهر تأثیر زیادی نخواهد گذاشت، اما می‌تواند منجر به پیچیده شدن فرآیند دسترسی خزندگان و ربات‌ها به سایت شما شود.

پیشنهاد: برای دستیابی به بزرگترین دیتابیس آنلاین از ربات‌های موجود در موتورهای جستجو، می‌توانید روی [این لینک](#) کلیک کنید.