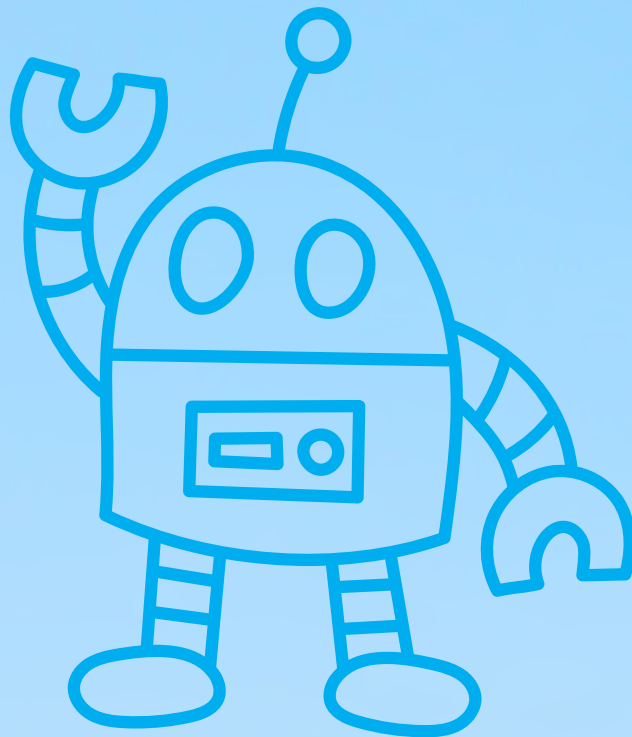


چيست و چه کاربردی دارد؟ Robots.txt



چيست و چه Robots.txt
کاربردی دارد؟



<https://iracode.com/wath-robots-txt/>



۰۹۱۲۴۸۰۴۵۳۵

ربات‌ها برنامه‌های خودکاری هستند که با طراحی وبسایت و طراحی اپلیکیشن در تعامل هستند. آن‌ها بدون دخالت انسان کارهای مشخصی را با سرعتی به مراتب بیشتر از انسان انجام دهند. یکی از این ربات‌ها، ربات خزنده وب است. این ربات وظیفه جستجو در سایت‌ها را برعهده دارد. فایل Robots.txt با فعالیتهای خزنده وب را مدیریت می‌کند و از اعمال کار اضافی به هاست جلوگیری کرده و به خزنده وب می‌گوید که چه صفحاتی را نمایش دهد و چه صفحاتی را نمایش ندهد. در ادامه این مقاله درباره این فایل بیشتر خواهید خواند.



Robots.txt چگونه کار می‌کند؟

جایگاه این فایل، مانند دیگر فایل‌های سایت در وب سرور است. در واقع اگر شما به آدرس کامل سایتی [Robots.txt](#) / را اضافه کنید صفحه مربوطه نمایش داده خواهد شد. این فایل لیستی از قوانین را در خود دارد؛ ولی به هر حال توانایی این را ندارد که ربات‌ها را ملزم به رعایت این قوانین کند. به عبارتی این دستورها مانند علایم راهنمایی و رانندگی هستند که به خودی خود، رانندگان را ملزم به رعایت قوانین مربوطه نمی‌کنند. در این میان فقط ربات‌های خوب هستند که از این قوانین تبعیت می‌کنند و ربات‌های بد این قوانین را نقض می‌کنند. ربات‌های خوب مانند خزنده وب، قبل از نمایش صفحات دیگر وبسایت، ابتدا به سراغ این فایل رفته و از قوانین پیروی می‌کنند. اما ربات‌های بد یا به این قوانین هیچ توجهی نمی‌کنند و یا سعی می‌کنند با استفاده از آن‌ها به صفحات ممنوعه دسترسی پیدا کنند. نمونه‌ای از فایل Robots.txt در زیر نمایش داده شده است:

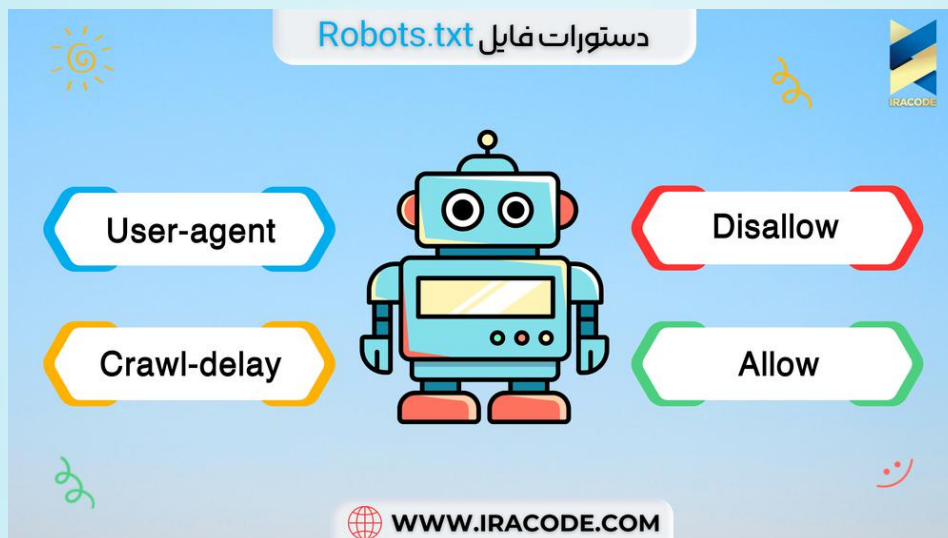
Robots.txt چیست و چه کاربردی دارد؟



```
User-agent: *
Disallow: /_esa
Disallow: /_mesa/
Disallow: /_xesa/
Disallow: /_csup/
Disallow: /_xsla/
Disallow: /_xcusp/
Disallow: /_xesa/
Disallow: /_xsla/
Disallow: /lp
Disallow: /feedback
Disallow: /langtest

Sitemap: https://www.cloudflare.com/sitemap.xml
Sitemap: https://www.cloudflare.com/fr-fr/sitemap.xml
Sitemap: https://www.cloudflare.com/de-de/sitemap.xml
Sitemap: https://www.cloudflare.com/es-es/sitemap.xml
Sitemap: https://www.cloudflare.com/pt-br/sitemap.xml
```

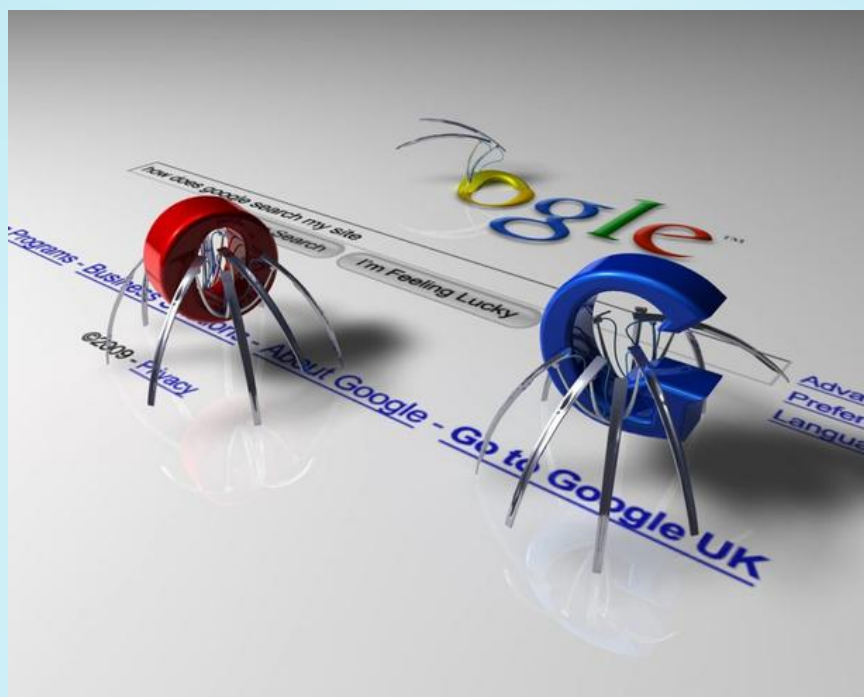
همانگونه که ملاحظه می‌کنید، سه عبارت User-agent, Disallow, Sitemap به کار برده شده است. البته علاوه بر این مولفه‌ها، مولفه دیگری به نام Crawl-delay و Allow هم وجود دارد که در اینجا آورده نشده است.



User-agent:

همه کاربران اعم از انسانی و ربات، دارای User agent یا عامل کاربر هستند. زمانی که کاربر انسانی از سیستم استفاده می‌کند، عامل کاربر شامل اطلاعاتی از قبیل نوع وب برورزر و نسخه سیستم عاملی که کاربر استفاده می‌کند می‌باشد. شایان ذکر است که عامل کاربر شامل هیچ‌گونه اطلاعات شخصی کاربر نمی‌باشد. اطلاعات مزبور به وبسایت کمک می‌کند تا محتوایی که با سیستم کاربر سازگاری دارند را نمایش دهد. هنگامی که کاربر از نوع ربات باشد عامل کاربر، به وبسایت کمک می‌کند تا مدیران سایتها نسبت به نوع رباتی که قصد دارد در سایت بخزد، آگاهی پیدا کنند. مدیران سایت می‌توانند با توجه به صلاحدید، دستورالعمل‌های خاصی را برای عامل کاربر رباتها تعریف کنند. به عنوان مثال، اگر مدیر سایتی بخواهد وبسایتش در صفحه نتایج گوگل نمایش داده شود، ولی در صفحه نتایج موتور جستجوگر بینگ نشان داده نشود، می‌تواند با استفاده از دستورالعمل‌های مربوطه این محدودیتها را اعمال کند.

دستور Disallow از دستورهای رایج فایل Robots.txt است که به رباتها دستور می‌دهد که به صفحه یا صفحاتی که بعد از این عبارت می‌آیند دسترسی نداشته باشد. البته این معنای مخفی کردن این صفحات نیست. بلکه این صفحات برای کاربران معمولی گوگل و بینگ مناسب نیستند. با این حال برخی کاربران این توانایی را دارند که به این صفحات دسترسی داشته باشند. چرا که این دسته از کاربران می‌دانند چگونه آنها را پیدا کنند. این دستور به روشها مختلف قابل استفاده است که تعدادی از آنها در تصویر بالا آورده شده است.



این دستور به موتور جستجو می‌فهماند که اجازه دسترسی به فایل یا صفحه خاصی را در دایرکتوری دارد. بعد از Allow لزوماً یک مسیر باید تعریف شود. در غیر این صورت، موتور جستجو آن را نادیده می‌گیرد. به مثال زیر توجه کنید:

* :User-agent

Allow: /media/terms-and-conditions.pdf

/Disallow: /media

در این مثال، اجازه دسترسی موتور جستجو به media داده نشده است، و تنها مورد مجاز در اینجا، media/terms-and-conditions.pdf است.

گاهی ممکن مدیر سایتی راهنمایی‌های متناقضی برای موتور جستجو بنویسد. به طوری که موتور جستجو نمی‌داند از کدام یک تبعیت کند. در چنین مواردی، موتور جستجو راهنمایی که کمترین

Robots.txt چیست و چه کاربردی دارد؟



محدودیت را ایجاد کند پیروی می‌کند. به این مثال دقت کنید:

User-agent: *

Allow: /directory

Disallow: *.html

در اینجا موتور جستجو نمی‌داند که آیا به صفحه `http://www.domain.com/directory.html` دسترسی داشته باشد یا خیر؟ با توجه به آنچه که گفتیم، موتور جستجو کمترین محدودیت را که دسترسی به این آدرس است را انتخاب می‌کند.

Crawl-delay:

این دستور فایل `Robots.txt` برای جلوگیری از خزش بیش از حد نوشته می‌شود. البته گوگل این دستور را پشتیبانی نمی‌کند. بنابراین وقت خود را صرف نوشتن این دستور برای خزنده گوگل نکنید. اما موتورهای جستجو مانند یاهو، بینگ و یاندکس از این دستور پشتیبانی می‌کنند. دقت داشته باشید که این دستور دقیقاً باید بعد از دستورات `Allow` و `Disallow` نوشته شود. به مثال زیر توجه کنید:

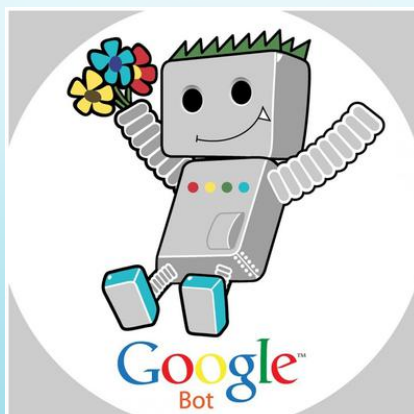
User-agent: BingBot

/Disallow: /private

Crawl-delay: 10

Robots.txt چه کاربردهایی دارد؟

- بستن صفحات غیر عمومی: گاهی ممکن است بخواهید برخی از صفحات سایتتان به هر دلیلی در معرض بازدید کاربران قرار نگیرد. در اینجا با اعمال دستور `Disallow` می‌توانید آن‌ها را از دید کاربران خارج کنید.
- بهینه کردن بودجه خزش سایت: در صورتی که تمام صفحات سایت شما قابلیت نمایش دارند، مشکل از بودجه خزش سایت شماست. با استفاده از این فایل، صفحات کم‌اهمیت را بلاک کرده و گوگل وقت خود را صرف باز کردن صفحات ارزشمند می‌کند.



آیا سایت من به Robots.txt نیاز دارد؟

اینکه سایت شما به این فایل نیاز دارد یا خیر به هدف شما از نمایش صفحات سایت بستگی دارد. همانگونه که گفتیم، یکی از کاربردهای این فایل این است که برخی از صفحات در معرض بازدید کاربران قرار نگیرد. بنابراین اگر چنین قصدی دارید و یا سایتی ساخته‌اید ولی هنوز نمی‌خواهید محتوای آن منتشر شود، ساخت این فایل برای شما مفید خواهد بود. اما اگر سایتی ساده دارید و چیزی برای پنهان کردن ندارید، به فایل Robots.txt نیاز نخواهید داشت. بنابراین استفاده نکردن از این فایل هم ممکن است مشکل خاصی ایجاد نکند. اما برای سایت‌های بزرگ، به سئو سایت کمک می‌کند.

سخن آخر

اگرچه که فایل Robot.txt به سئو سایت شما کمک می‌کند ولی باید بدانید که در صورت عدم استفاده صحیح از این ابزار و کدنویسی‌های نادرست، ممکن است به سایت خود ضربه بزنید. به عنوان مثال ممکن است با وارد کردن داده‌های نادرست، بخشی زیادی از سایت را از دید کاربران خارج کنید. بنابراین بهتر است از یک متخصص سئو برای نوشتن دستورات این فایل کمک بگیرید.