

Bad Bots y la pesadilla del tráfico. Htaccess en Apache 2.4

Cada vez esta pero el tema. Un ejercito de web scrappers, de personas dedicadas a vivir de crear contenido falso, indexable, o de robar imágenes, pulula por la red.

Una de las mejores formas de acabar con ellos es denegarles el acceso, en nuestro fichero `.htaccess`

.htaccess anti bad bots

La cuesgtión es añadir la lista de robots no deseados en nuestro fichero `.htaccce` usando para ello las directivas `setenvif`

```
# Start Bad Bot Prevention
<IfModule mod_setenvif.c>
# SetEnvIfNoCase User-Agent ^$ bad_bot
SetEnvIfNoCase User-Agent "^12soso.*" bad_bot
SetEnvIfNoCase User-Agent "^192.comAgent.*" bad_bot
SetEnvIfNoCase User-Agent "^1Noonbot.*" bad_bot
...

<Limit GET POST PUT>
  Order Allow,Deny
  Allow from all
  Deny from env=bad_bot
</Limit>
</IfModule>
# End Bad Bot Prevention
```

- `<IfModule mod_setenvif.c>` Esta directiva comprueba si el módulo `mod_setenvif` está habilitado. Si lo está, se ejecuta el código dentro de este bloque.
- `SetEnvIfNoCase User-Agent "^12soso.*" bad_bot`: Esta directiva establece una variable de entorno llamada `bad_bot` si el User-Agent comienza con "12soso".
- `<Limit GET POST PUT>`: Esta directiva limita las reglas dentro de este bloque a los métodos HTTP especificados (GET, POST y PUT).
- `Order Allow,Deny`: Define el orden en el que se aplican las reglas de acceso. Primero se aplican las reglas `Allow` y luego las reglas `Deny`.
- `Allow from all`: Permite el acceso a todos por defecto.
- `Deny from env=bad_bot`: Deniega el acceso a cualquier solicitud que tenga la variable de entorno `bad_bot` establecida.
- los `</` cierra la directiva correspondiente

La lista

Hay muchas, pero [esta lista](#) es un buen punto de partida y se actualiza regularmente. Incluso puedes mantenerla con algun pequeño script.

Aviso

Esta documentación y su contenido, no implica que funcione en tu caso o determinados casos. También implica que tienes conocimientos sobre lo que trata, y que en cualquier caso tienes copias de seguridad. El contenido el contenido se entrega, tal y como está, sin que ello implique ningún obligación ni responsabilidad por parte de [Castris](#)

Si necesitas soporte profesional puedes contratar con Castris [soporte profesional](#).

Revision #2

Created 16 February 2024 15:07:00 by Abkrim

Updated 16 May 2024 15:16:22 by Abkrim