

## Robots.txt file

Ово је фајл креиран у Notepad-у. Екстензија му је .txt. Служи да одреди које странице или фолдери неће бити индексирани и праћени. То је први фајл који спајери скенирају. Уколико нема забране за посећивање одређене странице, спајдер наставља са скенирањем дотичне странице. Снима се у root фолдер сајта на домену.

Корисни линк за више детаља: <https://varvy.com/robottxt.html>

Пример:

Allow full access

```
User-agent: *  
Disallow:
```

Block all access

```
User-agent: *  
Disallow: /
```

Block one folder

```
User-agent: *  
Disallow: /folder/
```

Block one file

```
User-agent: *  
Disallow: /file.html
```

Важно:

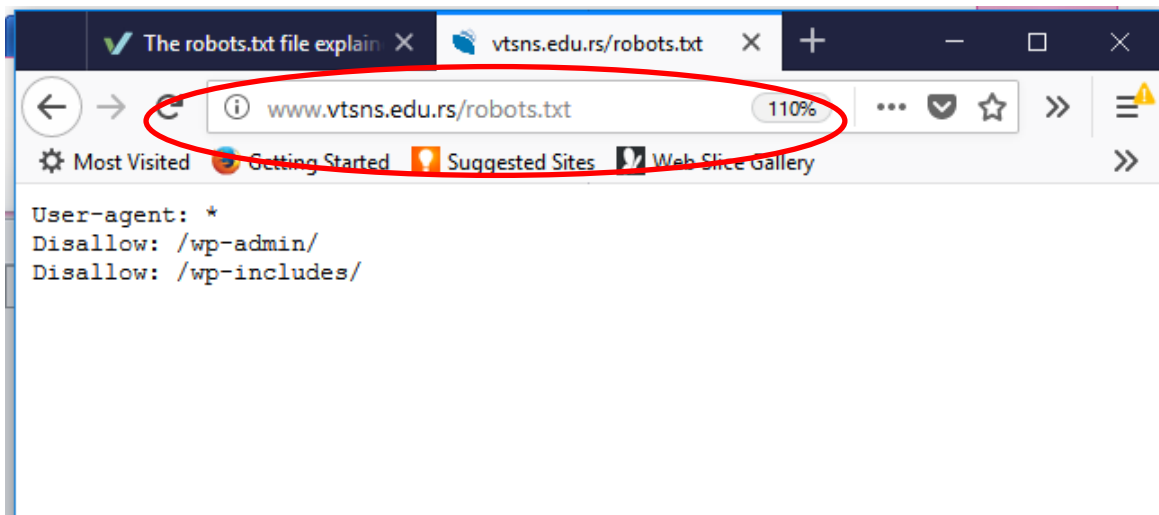
- Одредити да ли имате robots.txt фајл
- Уколико имате robots.txt фајл, проверите да ли вам блокира садржај који нисте желели да да буде блокиран
- Одредите да ли вам треба robots.txt фајл.

Веома лако се може проверити да ли имате robots.txt фајл.

Довољно је у веб читач укуцати следећу URL адресу:

```
www.yourwebsite.com/robots.txt
```

Пример проналажења robots.txt фајла ВТШ:



Блокирано је индексирање админ панела сајта.

Robots.txt фајл компаније win win

```
The robots.txt file explained an X vtsns.edu.rs/robots.txt winwin.rs/robots.txt
WinWin Shop doo Čačak (RS) https://www.winwin.rs/robots.txt
Most Visited Getting Started Suggested Sites Web Slice Gallery Висока техничка шко... Resizer

# Google Image Crawler Setup
User-agent: Googlebot-Image
Disallow:

# Crawlers Setup
User-agent: *

# Directories
Disallow: /app/
Disallow: /cgi-bin/
Disallow: /downloader/
Disallow: /errors/
Disallow: /includes/
Disallow: /lib/
Disallow: /media/captcha/
Disallow: /media/customer/
Disallow: /media/sales/
Disallow: /media/tmp/
Disallow: /pkginfo/
Disallow: /report/
Disallow: /scripts/
Disallow: /shell/
Disallow: /var/

# Paths (clean URLs)
Disallow: /index.php/
Disallow: /catalog/product_compare/
Disallow: /catalog/category/view/
Disallow: /catalog/product/view/
Disallow: /catalog/product/gallery/
Disallow: /catalogsearch/
Disallow: /control/
Disallow: /contacts/
Disallow: /customer/
Disallow: /customize/
Disallow: /newsletter/
Disallow: /poll/
Disallow: /review/
Disallow: /sendfriend/
Disallow: /tag/
Disallow: /wishlist/

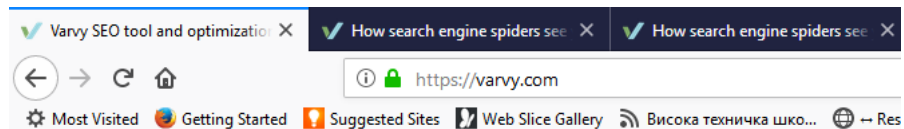
# Files
Disallow: /cron.php
Disallow: /cron.sh

# Paths (no clean URLs)
Disallow: /*.php$
Disallow: /*?SID=
```

Алтернатива је да нема овог фајла, као на примеру школе Ђура Јакшић из Зрењанина:



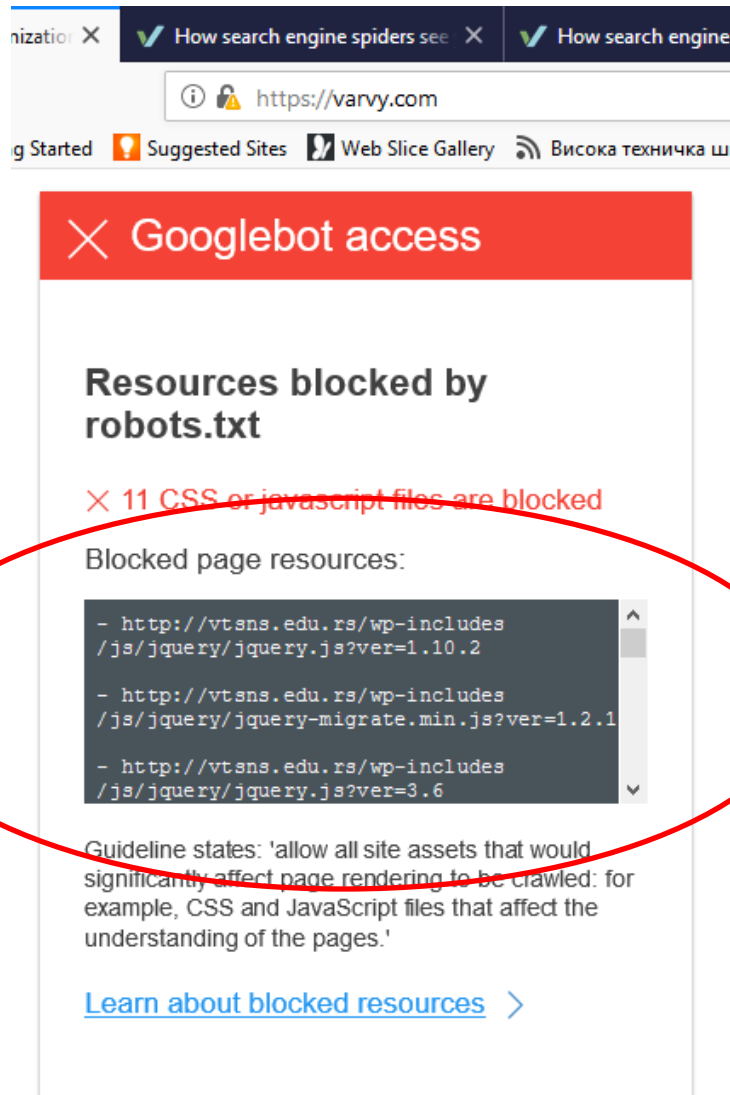
Алат који омогућава да проверимо да ли нам је блокиран важан садржај, <https://varvy.com/>



✓ seo speed mobile

## Varvy SEO tool

See how well a page follows the Google guidelines.



### Разлози постављања robots.txt фајла на сајт:

- Имамо садржај који желимо да сакријемо од робота
- Користимо плаћене линкове или рекламе које захтевају специјалне инструкције за роботе
- Сајт је у фази израде па не желимо да пре времена буде индексиран

### Разлози за не постављање robots.txt фајла на сајт:

- Нема могућности да погрешимо и да нам важан садржај буде блокиран
- Немамо фајлове које желимо сакрити од робота
- Не видимо разлог за поседовање robots.txt file-a

## User agent

Ово је наредба која нам омогућава да одредимо да наредбе које следе важе само за неки одређени робот. Уколико му је вредност \* - то значи да важи за све роботе.

```
User-agent: *
```

У примеру испод, наредбе које следе важиће само за Googlebot.

```
User-agent: Googlebot
```

Наредба Allow нам омогућава да дозволимо индексирање неког садржаја у фолдеру чије индексирање смо претходно онемогућили. На слици испод је блокиран садржај фолдера са фотографијама. Накнадно је дозвољено индексирање једне фотографије.

```
User-agent: *  
Disallow: /photos  
Allow: /photos/mycar.jpg
```