

Яндекс не индексирует 000webhost.com, настройка robots.txt

Стандарт исключений для роботов (robots.txt) — файл ограничения доступа к содержимому роботам на http-сервере. Файл должен находиться в корне сайта (то есть иметь путь относительно имени сайта **/robots.txt**). При наличии нескольких поддоменов файл должен располагаться в корневом каталоге каждого из них. Данный файл дополняет стандарт **Sitemaps**, который служит прямо противоположной цели: облегчать роботам доступ к содержимому.

Суть проблемы: **Яндекс робот проходит по сайту, но ничего не индексирует.** При этом, зайдя в [Яндекс Вебмастер](#), написано что сайт в очереди на индексацию и все, далее ничего не происходит уже 2 недели.

Есть такой замечательный плагин, называется [WP-log-robots](#). Очень рекомендую, ну просто надо, особенно на первое время. Пусть вас не смущает дата разработки - 2010 год. Плагин простой как 5 копеек, но эффективный. Он вам будет нужен для того, чтобы определить: когда (время) и какие роботы (поисковые системы) посещали ваш сайт, куда заходили, что индексировали, ну или по крайней мере пытались проиндексировать. Этот плагин в корне вашего сайта создает файл: **robots_log.txt**. Вы будете его открывать и видеть всех, кто заходит извне. Наиболее частый гость у меня **Googlebot**. Просто протоптал уже не тропинку, а конкретную такую колею.

Если посещения необходимых роботов имеются, но ни хрена не индексируется, тогда чешем репу. Мне, как среднестатистическому российскому распиздю, было все-равно происходит индексация или нет, но пора бы и честь знать. **Почему Яндекс не индексирует сайт?**

Ответ как всегда на поверхности, читайте [первоисточники](#).

Если быть кратким, то вот мой файл **robots.txt**, созданный по многочисленным рекомендациям особо грамотных товарищей:

```
User-agent: *
Allow: */uploads
Disallow: /cgi-bin
Disallow: /wp-
Disallow: */feed
Disallow: *?s=
Disallow: *?attachment_id=
```

Яндекс не индексирует 000webhost.com, настройка robots.txt

```
Disallow: /*?file_id=  
Disallow: /*?stats_author  
Disallow: /*?all_comments  
Disallow: /*?noindex  
Disallow: /id_date  
Disallow: /count.php?  
Disallow: /function-cat  
Disallow: /function-tag  
Disallow: /template-tags  
Disallow: /articles  
Disallow: /xmlrpc.php
```

```
Sitemap: http://it-engineer.ru/sitemap.xml  
Sitemap: http://it-engineer.ru/sitemap.xml.gz
```

```
Host: sysadminit.ru
```

Вот заметьте, не смотря на значение User-agent: * , эта звездочка прокатывает только для поисковика Google. Яндексу этот символ до одного места. В связи с этим, файл был быстренько переделан и залит обратно на сайт.

Выглядит мой новый файл **robots.txt** вот так:

```
User-agent: *  
Allow: */uploads  
Disallow: /cgi-bin  
Disallow: /wp-  
Disallow: */feed  
Disallow: /*?s=  
Disallow: /*?attachment_id=  
Disallow: /*?file_id=  
Disallow: /*?stats_author  
Disallow: /*?all_comments  
Disallow: /*?noindex  
Disallow: /id_date  
Disallow: /count.php?  
Disallow: /function-cat  
Disallow: /function-tag  
Disallow: /template-tags
```

Яндекс не индексирует 000webhost.com, настройка robots.txt

```
Disallow: /articles
Disallow: /xmlrpc.php
Sitemap: http://it-engineer.ru/sitemap.xml
Sitemap: http://it-engineer.ru/sitemap.xml.gz
```

Вышеуказанный участок файла – великолепно воспринимается абсолютно всеми роботами, за исключением яндекса.

```
User-agent: Yandex
Allow: */uploads
Disallow: /cgi-bin
Disallow: /wp-
Disallow: */feed
Disallow: *?s=
Disallow: *?attachment_id=
Disallow: *?file_id=
Disallow: *?stats_author
Disallow: *?all_comments
Disallow: *?noindex
Disallow: /id_date
Disallow: /count.php?
Disallow: /function-cat
Disallow: /function-tag
Disallow: /template-tags
Disallow: /articles
Disallow: /xmlrpc.php
Host: sysadminit.ru
```

Вот это 100% рабочий вариант файла **robots.txt** для блога **WordPress**.

Так что, пользуйтесь на здоровье.

Дополнение.

Ввиду того, что **YandexBot** — основной индексирующий робот поисковой службы Яндекса, работает с заграничными, а может и не только заграничными, хостингами через задницу, файл **robots.txt** будет периодически правиться, поэтому тех, кого

Яндекс не индексирует 000webhost.com, настройка robots.txt

интересует текущее состояние настроек, прошу смотреть [первоисточник](#).