Robots.txt

https://yandex.ru/support/webmaster/controlling-robot/robots-txt.xml

- В текстовом редакторе создайте файл с именем robots.txt и заполните его в соответствии с представленными ниже правилами.

- В роботе Яндекса используется сессионный принцип работы, на каждую сессию формируется определенный пул страниц, которые планирует загрузить робот.
- Сессия начинается с загрузки файла robots.txt. Если файл отсутствует, не является текстовым или на запрос робота возвращается HTTP-статус отличный от 200 OK, робот считает, что доступ к документам не ограничен.
- В файле robots.txt робот проверяет наличие записей, начинающихся с User-agent:, в них учитываются подстроки Yandex (регистр значения не имеет) или * . Если обнаружена строка User-agent: Yandex, директивы для User-agent: * не учитываются. Если строки User-agent: Yandex и User-agent: *отсутствуют, считается, что доступ роботу не ограничен.

- Директивы Disallow и Allow
- Чтобы запретить доступ робота к сайту или некоторым его разделам, используйте директиву Disallow.

```
User-agent: Yandex
Disallow: / # блокирует доступ ко всему сайту

User-agent: Yandex
Disallow: /cgi-bin # блокирует доступ к страницам,
                   # начинающимся с '/cgi-bin'
```

- Чтобы разрешить доступ робота к сайту или некоторым его разделам, используйте директиву Allow

```
User-agent: Yandex
Allow: /cgi-bin
Disallow: /
# запрещает скачивать все, кроме страниц
# начинающихся с '/cgi-bin'
```

- Использование спецсимволов * и $
- При указании путей директив Allow и Disallow можно использовать спецсимволы * и $, задавая, таким образом, определенные регулярные выражения. Спецсимвол * означает любую (в том числе пустую) последовательность символов.

Директивы Allow и Disallow без параметров

Если директивы не содержат параметры, учитывает данные следующим образом:

```
User-agent: Yandex
Disallow: # то же, что и Allow: /

User-agent: Yandex
Allow: # не учитывается роботом
```

Использование спецсимволов * и $

```
User-agent: Yandex
Disallow: /cgi-bin/*.aspx # запрещает '/cgi-bin/example.aspx'
                          # и '/cgi-bin/private/test.aspx'
Disallow: /*private # запрещает не только '/private',
                    # но и '/cgi-bin/private'
```

По умолчанию к концу каждого правила, описанного в файле `robots.txt`, приписывается спецсимвол *.
Пример:

```
User-agent: Yandex
Disallow: /cgi-bin* # блокирует доступ к страницам
                    # начинающимся с '/cgi-bin'
Disallow: /cgi-bin # то же самое
```

Чтобы отменить * на конце правила, можно использовать спецсимвол $, например:

```
User-agent: Yandex
Disallow: /example$ # запрещает '/example',
                    # но не запрещает '/example.html'
```

```
User-agent: Yandex
Disallow: /example # запрещает и '/example',
                   # и '/example.html'
```

Спецсимвол $ не запрещает указанный * на конце, то есть:

```
User-agent: Yandex
Disallow: /example$  # запрещает только '/example'
Disallow: /example*$ # так же, как 'Disallow: /example'
                     # запрещает и /example.html и /example
```

Директива Sitemap

Если вы используете описание структуры сайта с помощью файла Sitemap, укажите путь к файлу в качестве параметра директивы `Sitemap` (если файлов несколько, укажите все). Пример:

```
User-agent: Yandex
Allow: /
Sitemap: http://example.com/site_structure/my_sitemaps1.xml
Sitemap: http://example.com/site_structure/my_sitemaps2.xml
```

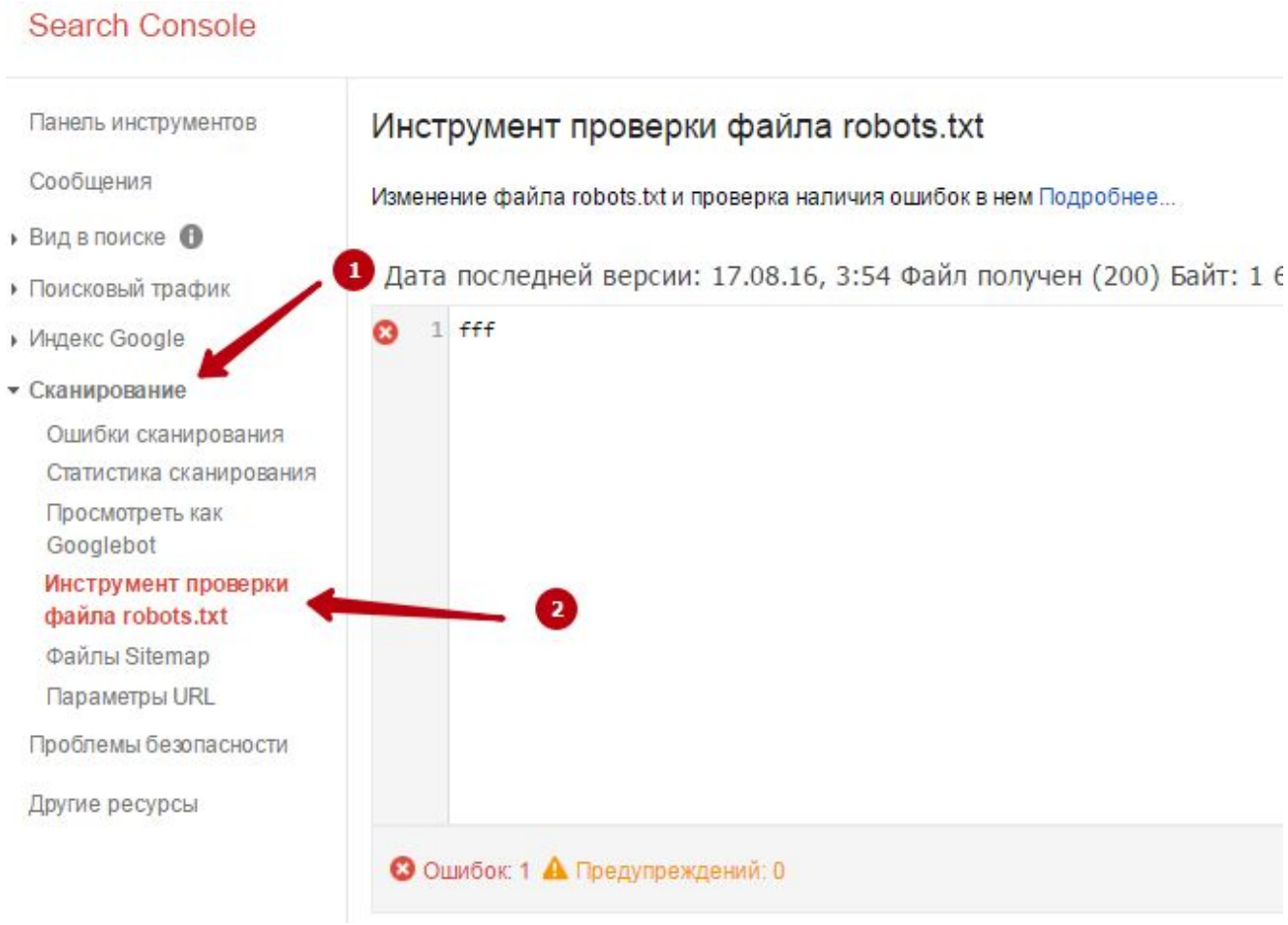
Робот запомнит путь к файлу, обработает данные и будет использовать результаты при последующем формировании сессий загрузки.

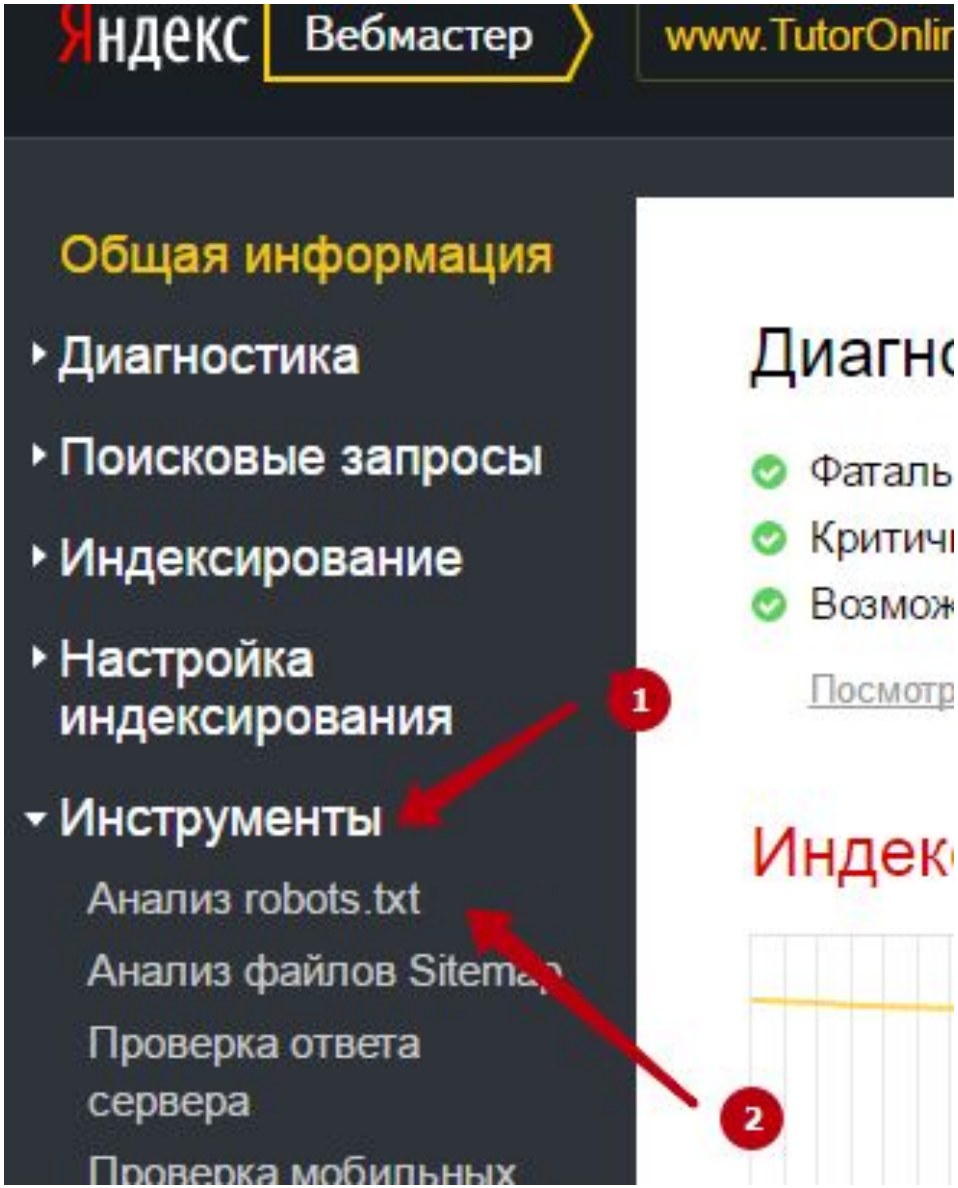
Директива Host

- Если у вашего сайта есть зеркала, специальный робот зеркальщик определит их и сформирует группу зеркал вашего сайта. В поиске будет участвовать только главное зеркало. Вы можете указать его для всех зеркал в файле robots.txt: имя главного зеркала должно быть значением директивы Host.
- Директива 'Host' не гарантирует выбор указанного главного зеркала, тем не менее, алгоритм при принятии решения учитывает ее с высоким приоритетом

```
#Если www.glavnoye-zerkalo.ru главное зеркало сайта, то  #robots.txt для всех сайтов из г
руппы зеркал выглядит так
User-Agent: *
Disallow: /forum
Disallow: /cgi-bin
Host: www.glavnoye-zerkalo.ru
```

Проверьте ваш файл в вебмастерах google и yandex




Яндекс
Вебмастер
www.TutorOnlin
Общая информация
Диагностика
Поисковые запросы
Индексирование
Настройка индексирования
Инструменты
Анализ robots.txt
Анализ файлов Sitema
Проверка ответа сервера
Проверка мобильных
Диагно
Фаталь
Критичн
Возмож
Посмотр
Индек
1
2