

攻防世界 web robots

原创

[_n19hT](#) 于 2020-02-27 11:14:08 发布 964 收藏

分类专栏: [# web](#) 文章标签: [搜索引擎](#) [网络](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: https://blog.csdn.net/weixin_43092232/article/details/104532082

版权



[web](#) 专栏收录该内容

13 篇文章 0 订阅

订阅专栏

终于见到了一个协议了!!!


题目描述: X老师上课讲了Robots协议, 小宁同学却上课打了瞌睡, 赶紧来教教小宁Robots协议是什么吧。

通过题目描述就知道这题要用到robots协议，所以百度一下。

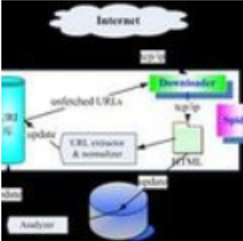
度 robots协议  **百度一下**

网页 资讯 视频 图片 知道 文库 贴吧 采购 地图 更多»

百度为您找到相关结果约11,200,000个


 搜索工具

[robots协议_百度百科](#) 



robots协议也叫robots.txt (统一小写) 是一种存放于网站根目录下的ASCII编码的文本文件, 它通常告诉网络搜索引擎的漫游器 (又称网络蜘蛛), 此网站中的哪些内容是不应被搜索引擎的漫游器获取的, 哪些是可以被漫游器获取的。因为一些系统中...

[简介](#) [原则](#) [功能](#) [位置](#) [产生](#) [影响](#) [搜索引擎](#) [更多>>](#)

<https://baike.baidu.com/> 



[Robots协议 - 简书](#)


2018年11月29日 - 好的网络爬虫,首先需要遵守Robots协议。Robots协议(也称为爬虫协议、机器人协议等)的全称是“网络爬虫排除标准”(Robots Exclusion Protocol)...

 [简书社区](#)  - [百度快照](#)

[robots协议文件的写法及语法属性解释 - white_HATmagic - CSDN博客](#)

2018年10月28日 - 当一个搜索蜘蛛访问一个站点时,它会首先检查该站点根目录下是否存在robots.txt,如果存在,搜索机器人就会按照该文件中的内容来确定访问的范围;如果该文...

 [CSDN技术社区](#)  - [百度快照](#)

[如何利用百度查看网站的Robots协议_百度经验](#) 

条评价
制指令 (限制搜索引擎抓取)

- 1 在搜索框里面随便输入你想搜索的信息如...
- 2 鼠标移到了解详情, 左键单击 --> 进入...
- 3 可以输入你想要了解的网站的网址我们在...
- 4 可以看到下面的文本框出现了很多的脚本...

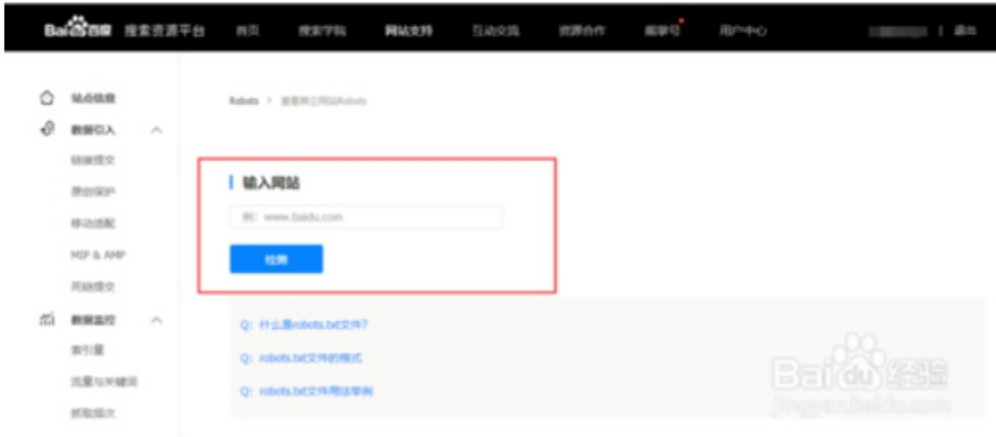
[显示全部](#) 

https://blog.csdn.net/weixin_43092232

robots协议也叫robots.txt (统一小写) 是一种存放于网站根目录下的ASCII编码的文本文件, 它通常告诉网络搜索引擎的漫游器 (又称网络蜘蛛), 此网站中的哪些内容是不应被搜索引擎的漫游器获取的, 哪些是可以被漫游器获取的。因为一些系统中的URL是大小写敏感的, 所以robots.txt的文件名应统一为小写。robots.txt应放置于网站的根目录下。如果想单独定义搜索引擎的漫游器访问子目录时的行为, 那么可以将自定的设置合并到根目录下的robots.txt, 或者使用robots元数据 (Metadata, 又称元数据)。

然后下面一个百度经验说的是怎么查看网站的robots.txt

7 方法二：
浏览器上直接输入：<https://ziyuan.baidu.com/robots/index>
也可以进入



https://blog.csdn.net/weixin_43092232

输入网站

检测

1次重试中, 请稍等.....

Q: [什么是robots.txt文件?](#)

Q: [robots.txt文件的格式](#)

Q: [robots.txt文件用法举例](#)


https://blog.csdn.net/weixin_43092232

但是我把这个地址放进去是找不到robot.txt的
然后我又试了把在URL后面加上/robots.txt, 即

<http://111.198.29.45:38915/robots.txt>

← → ↻ ⓘ 不安全 | 111.198.29.45:38915/robots.txt

```
User-agent: *  
Disallow:  
Disallow: flag_1s_h3re.php
```



https://blog.csdn.net/weixin_43092232

我把这个当成字符串交上去结果发现不对，想了想又把URL的后缀改成了这个。
成功找到flag!

← → ↻ ⓘ 不安全 | 111.198.29.45:38915/flag_1s_h3re.php

cyberpeace{efa88389b3358a51ad4518e34b42ce21}

cyberpeace{efa88389b3358a51ad4518e34b42ce21}

渐渐找到web的感觉