

攻防世界web Training-WWW-Robots writeup

原创

[Sprint#51264](#) 于 2020-07-08 18:03:25 发布 83 收藏

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/qq_45837896/article/details/107211656

版权

In this little training challenge, you are going to learn about the [Robots exclusion standard](#).

The robots.txt file is used by web crawlers to check if they are allowed to crawl and index your website or only parts of it.

Sometimes these files reveal the directory structure instead protecting the content from being crawled.

Enjoy!

https://blog.csdn.net/qq_45837896

题目描述见上；

根据题目要求我们学习robots协议相关知识。

robots.txt:

该协议将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。

所以使用搜索引擎的时候通过散布spider进行获取网站内容时，spider会先检索是否有robots.txt文件，如果有，那么spider将会根据该文件中的限制进行检索。

于是我们在在线环境url后输入/robots.txt查看该文件

```
User-agent: *  
Disallow: /f10g.php
```

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

https://blog.csdn.net/qq_45837896

发现它不允许/f10g.php这个文件被检索，那么我们在地址栏中输入这个文件名进行进入。

220.249.52.133:46430/fl0g.php

cyberpeace{fe6071433c74a0b813af50c517e03819}

拿到flag。