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

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

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。

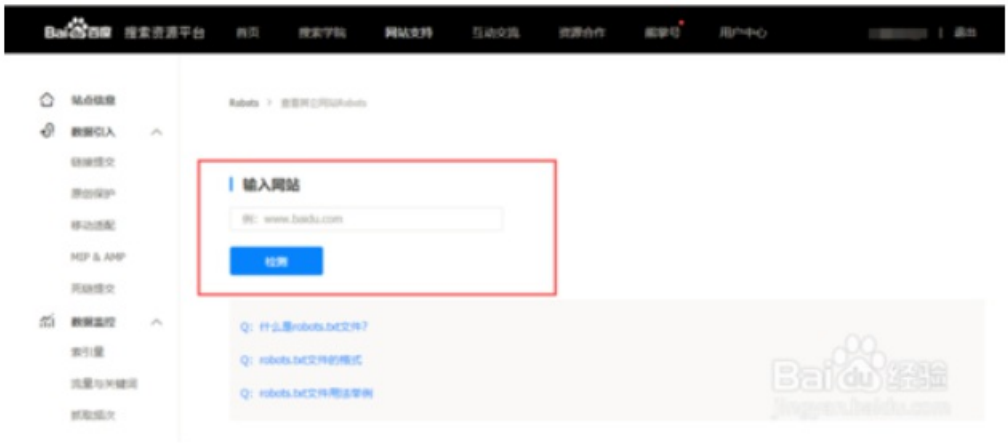
> robots协议也叫robots.txt（统一小写）是一种存放于网站根目录下的ASCII编码的文本文件，它通常告诉网络搜索引擎的漫游器（又称网络蜘蛛），此网站中的哪些内容是不应被搜索引擎的漫游器获取的，哪些是可以被漫游器获取的。因为一些系统中的URL是大小写敏感的，所以robots.txt的文件名应统一为小写。robots.txt应放置于网站的根目录下。如果想单独定义搜索引擎的漫游器访问子目录时的行为，那么可以将自定的设置合并到根目录下的robots.txt，或者使用robots元数据（Metadata，又称元数据）。

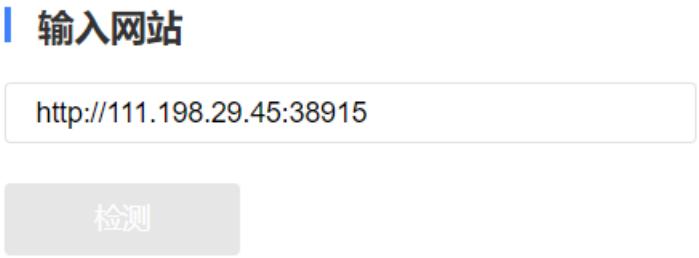
然后下面一个百度经验说的是怎么查看网站的robots.txt

7 **方法二：**

浏览器上直接输入：https://ziyuan.baidu.com/robots/index

也可以进入





Q：什么是robots.txt文件?

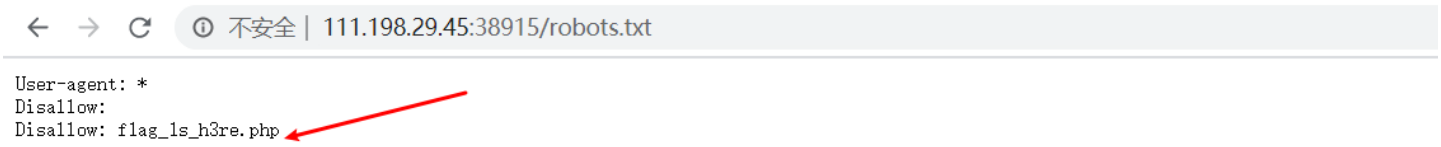
Q：robots.txt文件的格式

Q：robots.txt文件用法举例

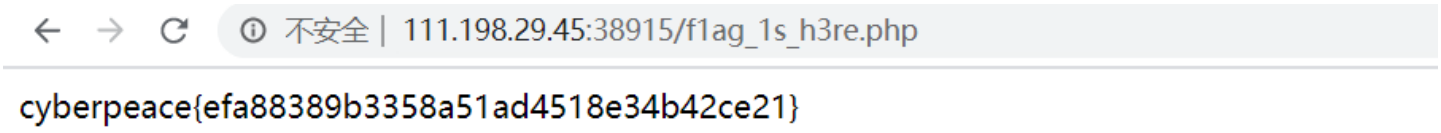
但是我把这个地址放进去是找不到robot.txt的

然后我又试了把在URL后面加上/robots.txt，即

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



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！



cyberpeace{efa88389b3358a51ad4518e34b42ce21}

渐渐找到web的感觉