

## Training-WWW-Robots

### 解题思路:

- 开头一大段英语, 大概就是告诉我们: 有个叫 `robots.txt` 的文档, 这个文档记录了 **不应该被搜索引擎的漫游器获取的东西**。

- 在查阅资料后, 知道它一般放在根目录下, 搜索引擎搜索这个网站时, 首先访问这个 `robots.txt`, 如果它存在, 那就按照它上面写的, 该不碰的页面就不碰, 如果它不存在, 那搜索引擎就可以查阅所有没有口令保护的页面。

当然, 这只是一种约定俗成的东西, 并没有实质上的防御能力。

- 知道了之后, 由于 `robots.txt` 在根目录, 所以我们直接在 `url` 后面加一个 `robots.txt`, 构成一个新的 `url`: `111.198.29.45:39110/robots.txt`

- 发送这个请求, 服务器就会向我们返回它 `robots.txt` 里面的内容:

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

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

`disallow` 指的就是不允许访问, 但整个页面总共就这里面出现了一个 `.php` 文件, 虽然知道这样不太道德, 但我们为了解题, 只好访问一下它啦。

- 再次构造请求: `http://111.198.29.45:39110//f10g.php`, 发送请求, 服务器就给我们回传 `flag` 啦!