## 实验4 机器人协议

|  |  |  |  |
| --- | --- | --- | --- |
| 评语 | 作业评语 | 得分 | 成绩 |

### 1 实验目的

* 熟悉DNS缓存的处理。
* 熟悉机器人协议的获取方式。
* 熟悉机器人协议的内容。

### 2 实验准备

* 安装好PyCharm集成开发环境并配置好
* 了解Python程序基本结构
* 掌握如何实现DNS的缓存处理
* 掌握如何获取机器人协议
* 了解机器人协议的内容

### 3 实验内容

以下为网站的访问资源，请根据要求设定机器人协议的内容，并编写程序验证协议内容的正确性：

/rebots.txt

/index.html

/private/index.html

/friend/index.html

/other/index.html

/public/index.html

要求：

1. 当爬虫名称为friend时，可以爬起根下的index.html及friend、public下的内容
2. 当爬虫名称为classmate时，可以爬取根下的index.html及other、public下的内容
3. 当爬虫名称为你的学号时，可以爬取所有的内容

**4 实验结果代码**

【机器人协议的内容】

【验证的代码】