课程名称 python实验

实验成绩

指导教师



实 验 报 告

院系

班级   
学号

姓名

日期

# Python语言实验报告



## 实验八: python爬虫

## 一 实验目的

1)robots协议

2)tomcat的安装

3)搭建单机开发环境:最简单的网页

4)创建自己的robots.txt

## 二 实验环境

计算机：CPU四核i7 6700处理器；内存8G； SATA硬盘2TB硬盘； Intel芯片组

主板；集成声卡、千兆网卡、显卡； 20寸液晶显示器。

环境：windows7 64位、python3.5.3、pycharm教育版

## 三 实验内容

### 3.1 实验内容一

#### 3.1.1实验题目

**1 打开豆瓣网页：**<https://www.douban.com/>

1 读取 robots.txt文件，截图填入报告

2 从禁止爬虫爬取的子目录里面选择一个填入报告，例如：

User-agent: \*

Disallow: /subject\_search

表示对于任何一种爬虫，豆瓣网站上/subject\_search路径下的资源不允许爬取

#### 3.1.2实验步骤

#### 3.1.3实验结果(截图)

1.



2.

User-agent: Mediapartners-Google

Disallow: /j/

不允许谷歌的爬虫爬取J路径下的资源

#### 3.1.4实验结论

结合自身理解，写出实验结果分析、实验收获，知识点掌握程度等。

### 3.2 实验内容二

#### 3.2.1实验题目

**2 安装apache tomcat**

完成安装后打开:http://127.0.0.1:8080,截图,填入实验报告

#### 3.2.2实验步骤

#### 3.2.3 实验结果(截图)



#### 3.2.4 实验结论

结合自身理解，写出实验结果分析、实验收获等。

### 3.3 实验内容三

#### 3.3.1实验题目

### 3 编写一个简单的网页

在tomcat的webapps目录下新建python目录,在里面创建index.html文件,输入以下内容:



写好以后,使用浏览器打开:

调整浏览器窗口大小,把浏览器截图填入实验报告,注意必须截取浏览器显示网址的部分

#### 3.3.2实验步骤

#### 3.3.3 实验结果(截图)



#### 3.3.4 实验结论

结合自身理解，写出实验结果分析、实验收获等。

### 3.4 实验内容四

#### 3.4.1实验题目

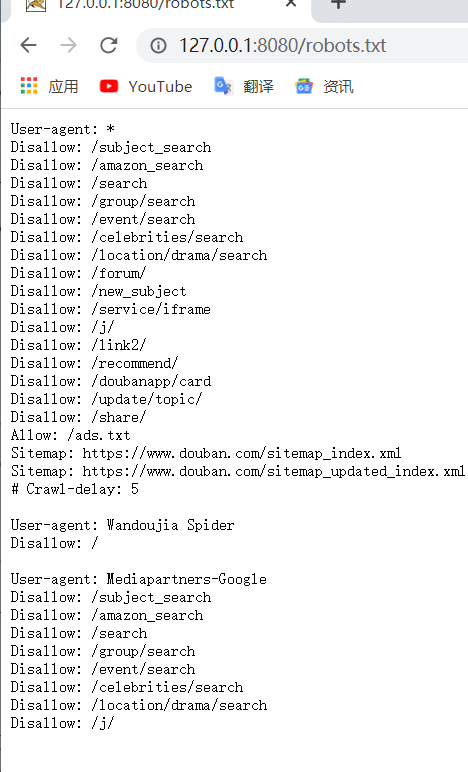
### 4 仿照豆瓣的爬虫协议，创建自己的robots协议，使用浏览器访问，截图，填入报告。

注意思考：robots.txt文件应该放在哪一个文件夹下

#### 3.4.2实验步骤

放在webapps\ROOT文件夹下

#### 3.4.3 实验结果(截图)



#### 3.4.4 实验结论

结合自身理解，写出实验结果分析、实验收获等。

## 四 实验感想

对于本次实验课，收获了什么，发现了那些不足，能力是否得到提高等内容