

计算机与信息学院课程设计报告

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 课程名称： | | Python程序设计 | | | | |
| 课程设计名称： | | **毕业生租房信息查询** | | | | |
| 实验人员： | **人员** | | **学号** | | **姓名** | | |
| 组长 | | 20122003078 | | 汪玉叶 | | |
| 组员1 | | 20122003020 | | 范垄垄 | | |
| 组员2 | |  | |  | | |
| 班级 | | 2020级物联网工程（专升本） | | | | |
| 实验日期： | 2021-5-31 | | | | | | |
|  |  | | | | | | |
| 课程设计评价： |  | | | | | | |
|  | 设计成绩： | |  | 评定日期： | |  | |
|  | 指导教师： | |  | | | | |

目录

[一、需求分析 3](#_Toc74932367)

[二、总体设计 3](#_Toc74932368)

[三、详细设计 5](#_Toc74932369)

[四、总结 8](#_Toc74932373)

[五、小组分工 8](#_Toc74932374)

## 摘要

许多大学生毕业之后都面临着一个除了找工作之外同样让人头疼的问题，就是租房。对于我们班的同学也是，马上也要面临毕业了。租房在目前来看已经成为一种趋势，但是在陌生的环境中，对地方不熟悉很容易导致我们被骗，或是被中介赚取高额差价。除此之外，我们每天上班的来回通勤时间，房子是否太贵超过我们的负担，以及房子本身的软硬件设施都是我们要考虑的问题。这样一来，不仅耗费我们大量的时间去找房，还不一定能找到令我们满意的房源。此项目灵感就来源于此，为此，我们组设计了一个毕业生租房信息查询的爬虫小工具，在一定程度上能够帮助我们解决一些问题。

**关键词**：毕业生、租房、爬虫

## 一、需求分析

许多大学生毕业之后都面临着一个除了找工作之外同样让人头疼的问题，就是租房。对于我们班的同学也是，马上也要面临毕业了。租房在目前来看已经成为一种趋势，但是在陌生的环境中，对地方不熟悉很容易导致我们被骗，或是被中介赚取高额差价。除此之外，我们每天上班的来回通勤时间，房子是否太贵超过我们的负担，以及房子本身的软硬件设施都是我们要考虑的问题。这样一来，不仅耗费我们大量的时间去找房，还不一定能找到令我们满意的房源。此项目灵感就来源于此，为此，我们组设计了一个毕业生租房信息查询的小工具，在一定程度上能够帮助我们解决一些问题。

总结一下大学生租房难的症结：

1. 没钱
2. 小中介发布的价位一般都是假的，会浪费你很多时间。
3. 如果是多人，得同时考虑多人的通勤时间。
4. 对交通路线不熟悉以致于选择面窄。

## 二、总体设计

实验总体分为两大部分，一是分析页面，爬取房源信息。二是将高德地图可视化。实验中会用到三个文件：crawl.py、rent.csv与index.html，其中crawl.py 是一个非常简单的爬取网页的脚本。rent.csv 是由 crawl.py 生成的房源文件。index.html 是最重要的显示地图的部分。实验流程如下图2.1所示：

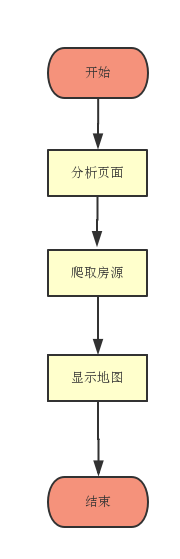


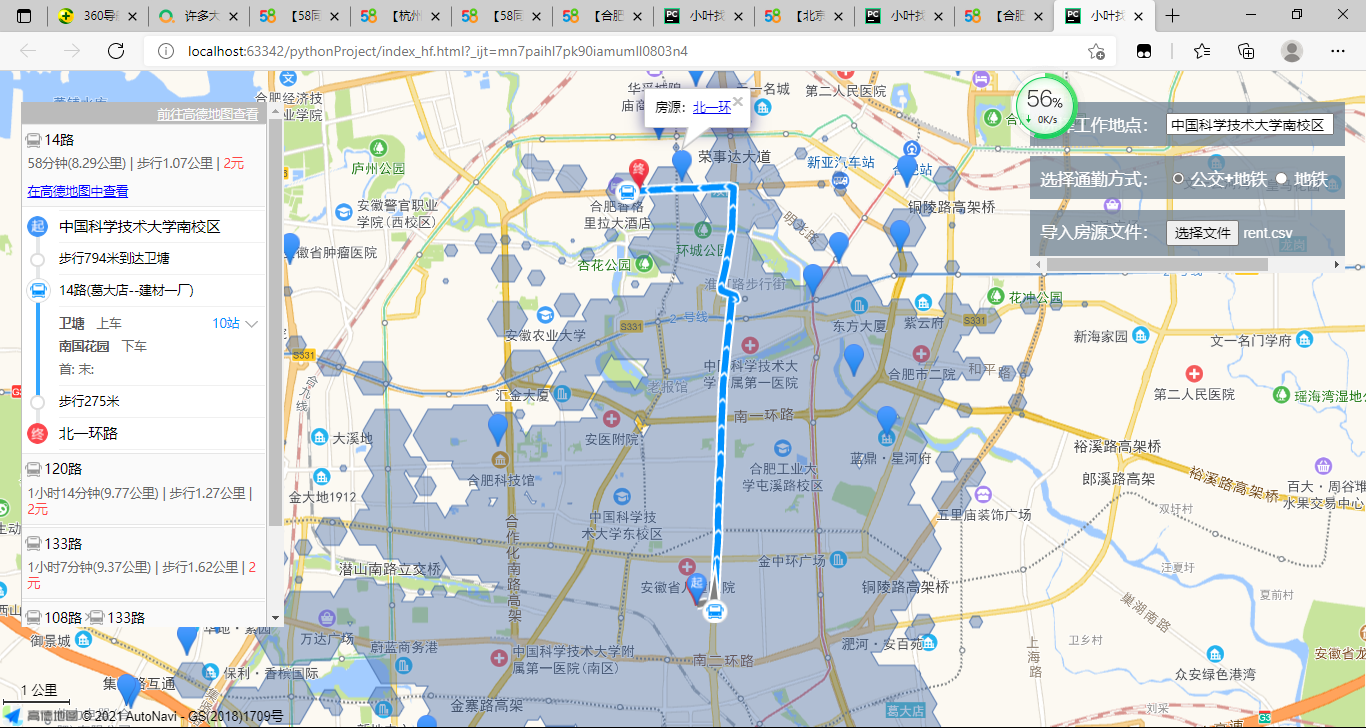
图 2.1

实验主界面：



图 2.2

最终效果图：



如图，划出来的大片蓝色色块就是距离工作地点一小时车程内的区域。蓝色的标记就是房源，点击后会自动给出路径规划和房源地址。红色标记（不是"终"）是工作地点。工作地点的输入框有自动补完的功能。

## 三、详细设计

主要用到的技术有pyqt5,requests、BeautifulSoup、csv及高德地图API。

requests是一个对使用者非常友好的http库，看一遍[Quickstart](http://www.python-requests.org/en/master/user/quickstart/" \t "https://www.lanqiao.cn/courses/599/learning/_blank)就能基本掌握它的使用。

Beautiful Soup是一个用来解析html或者xml文件的库，支持元素选择器，使用起来也非常方便。

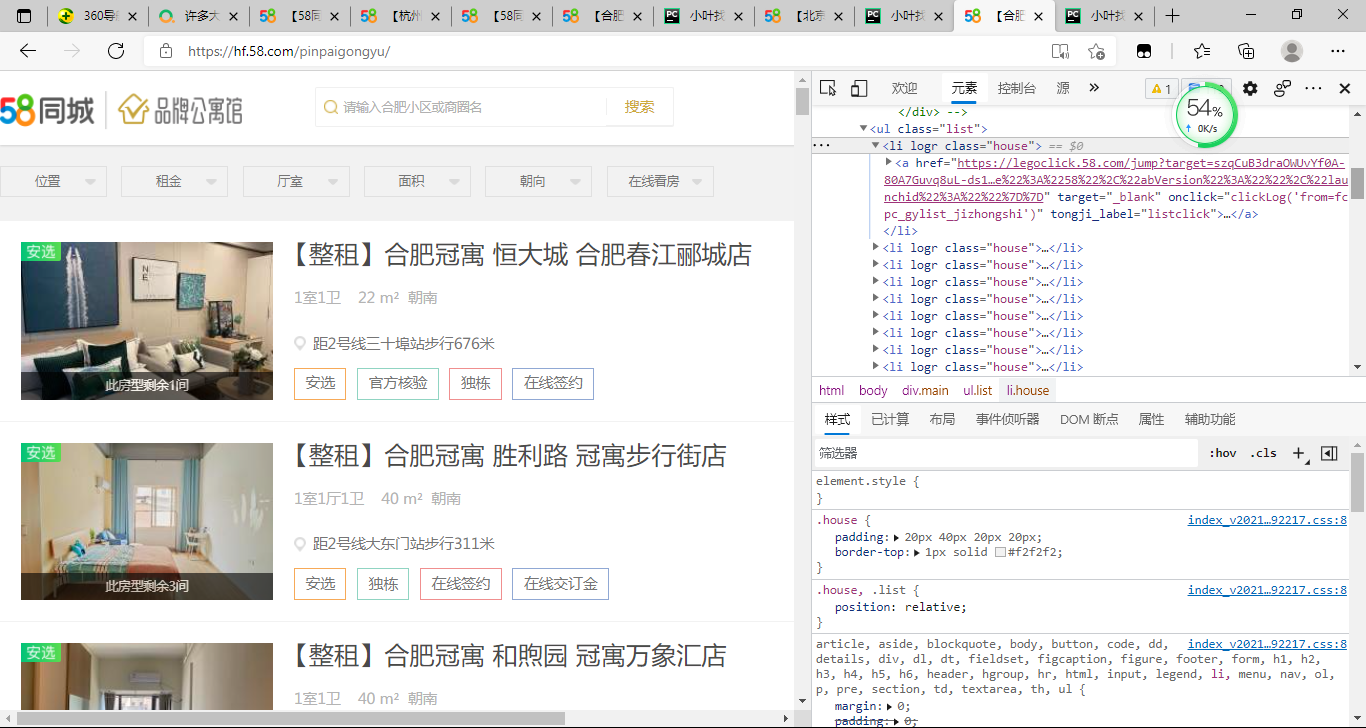
csv一般用作表格文件，直接用文本编辑器打开也可读，行与行之间就是换行来隔开，列与列之间就是用逗号（也可指定其它字符）隔开，Python 标准库中的csv库就是用来读写csv文件的。

注册高德的开发者用户，创建应用就能得到key。这部分借鉴高德地图示例中心及参考手册。

### 分析页面

首先需要分析我们想爬取的页面





比如上图所示页面，我们需要右击检查元素，我们发现所有的租房信息都是一个li标签，并且都位于ul标签下，当我们往下移的时候，相应的标签也会越来越多，所以我们便可以利用有无list类来判定是否为租房请求。当然，这是在没有任何条件获取到的全部信息，但很多信息我们是不会看的，既然有租房的意愿，那么我们就有一个大致的心理承受价格，因此我们可以来分析一下该网站的 url 构造形式，大致了解了它的路径规则：/pingpaigongyu/pn/{page}/minprice={min\_rent}\_{max\_rent}。

### 爬取房源

爬取房源就可以用我们上课学的爬虫来实现了。主要代码如下：

#!/usr/bin/env python3  
from bs4 import BeautifulSoup # 网页解析模块  
import requests # 网络请求模块  
import csv # csv 文件模块  
import time  
import lxml  
# 网址  
url = "https://hf.58.com/pinpaigongyu/pn/{page}/?minprice=1000-2000"  
  
# 初始化页码  
page = 0  
# 打开rent.csv文件，如果没有就创建一个，并设置写入模式  
csv\_file = open("rent.csv","w",encoding="utf-8")  
# 创建writer对象  
csv\_writer = csv.writer(csv\_file, delimiter=',')  
# 循环所有页面  
while True:  
 page += 1  
 print("fetch: ", url.format(page=page))  
 time.sleep(1)  
 # 抓取目标页面  
 response = requests.get(url.format(page=page))  
 # 设置编码模式  
 response.encoding = 'utf-8'  
 # 创建一个BeautifulSoup对象，获取页面正文  
 html = BeautifulSoup(response.text,features="lxml")  
 # 获取当前页面的房子信息  
 house\_list = html.select(".list > li")  
  
 # 循环在读不到新的房源时结束  
 if not house\_list:  
 break  
 # 房子信息  
 for house in house\_list:  
 # 获取房子标题  
 house\_title = house.select("h2")[0].string  
 # 获取房子链接地址  
 house\_url = house.select("a")[0]["href"]  
 # 对标题进行分隔  
 house\_info\_list = house\_title.split()  
 # 如果第二列是公寓名则取第一列作为地址  
 if "公寓" in house\_info\_list[1] or "青年社区" in house\_info\_list[1]:  
 house\_location = house\_info\_list[0]  
 else:  
 house\_location = house\_info\_list[1]  
  
 house\_money = house.select(".money")[0].select("b")[0].string  
 # 写入一行数据  
 csv\_writer.writerow([house\_title, house\_location, house\_money, house\_url])  
 # 关闭文件  
csv\_file.close()

### 显示地图

这部分主要用高德地图Javascript API来实现。这部分有个任务需要提前主备就是要自己注册成为高德地图的开发者，申请一个key。如下代码所示：

<script src="http://webapi.amap.com/maps?v=1.3&key=36aa77404a1c2f2fc922a569eaa49422&plugin=AMap.ArrivalRange,AMap.Scale,AMap.Geocoder,AMap.Transfer,AMap.Autocomplete">

</script>

载入编写代码时要用到的 API 插件：

<script src="http://webapi.amap.com/maps?v=1.3&key=36aa77404a1c2f2fc922a569eaa49422&plugin=AMap.ArrivalRange,AMap.Scale,AMap.Geocoder,AMap.Transfer,AMap.Autocomplete">

</script>

在页面的head标签中加入：

<meta

http-equiv="Content-Security-Policy"

content="upgrade-insecure-requests"

/>

意思是自动将 http 的不安全请求升级为 https。

## 四、总结

互联网时代，信息唾手可得，但是呢，我花大把时间看到的都不是我想要的，即使在看个人房源还是得自己排除一堆中介。我不是常常出门所以对交通不熟，把这些信息都查一遍又得花大量的时间。还好现在有了各种好的平台与开放接口，只需要你有一点点编程技能和一个想要实现什么东西的想法，一切就都不一样了。

## 五、小组分工

组长汪玉叶：负责编写爬虫、调用高德地图API、演示作品

组员范垄垄：负责设计和编写软件界面、撰写实验报告和PPT