**计算机科学与工程学院 实验报告**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **实验课程名称** | | **Python编程与数据分析project2** | | | **实验总成绩** |  |
| **专业** |  | | **班级** |  | **指导教师签字** |  |
| **学号** |  | | **姓名** |  | **实验报告批改时间** |  |
|  | | | | | | |
| 1. Project2思路：   1.实验题目中要求复用project1中的一些函数与方法，我在我的project中主要复 用了：   1. project1中的fuzzysearch()函数，用来对输入的城市进行模糊搜索 2. 日志模块 3. Choose\_city\_num()函数，对输入的城市进行在城市文件中的选择。   2.选用pycharm + pyqt5进行GUI设计，整体采用软件开发MVC架构，实现逻辑代码与前端界面相分离。之所以选用pyqt5集pyside2主要是因为我对Qt比较熟悉。  3.对于爬虫部分我使用request中的get方法进行数据收集，再使用parsel中的Selector方法进行数据处理，将网页<li> …</li>中的数据保存在字典中；再调用csv库中的writerow方法将字典中的数据保存在csv文件中；再从csv文件中读取数据保存到数据库中。对于爬虫，我使用的方法需要使用”User-Agent”用户代理，但是不同网页的用户代理不一样，为了解决自动爬取这个问题，我结合计算机网络课程所学知识，在“链家”网站上注册了账号并登录，之后”User-Agent”用户代理对于同一网站的不同网页就都一样了，由此实现了切换网页的自动爬取。  4.对于保存到数据库，我选用的数据库是《数据库原理》课程行所学的postgresql数据库。在程序中调用psycopg2库中的方法：先调用connect方法进行数据库的连接；再建立游标进行数据库建表、插入、查询等语句的执行。  5.对于最终查询数据的展示，我采用了三种方式：（1）在csv文件中展示；（2）在GUI界面中的qtablewidget中以表格的形式展示；（3）用DataFram类型的数据结构直接从数据库中读出所有的数据并在textEdit中进行展示。     1. 界面展示：   （1）    这是运行程序之后的主界面，右上角有两个按钮，现在显示的分别是“未登录”和“立即注册”，中间长条的是搜索框，由于现在没有登录所以不能搜索。  （2）  （注册界面） （登录界面）  登录界面和注册界面中有填写验证码的步骤，点击验证码的按钮还可以更换验证码，注册的用户名和密码会存入数据库中，等下一个用户注册的时候会用用户名取和数据库中已有的用户名进行比较，并会有相应的弹窗进行提示。    （数据库中存储用户信息的表userinfo）    （提示信息之一）  （4）登录后右上角的图标和字样会发生变化，并且中间输入框中的提示信息也会发生变化。  （5）输入“京”之后按回车会出现模糊搜索的搜索结果，搜索结果在一个下拉框中展示，由于电脑截图的问题在此无法展示。    （6）按回车键选中想搜索的城市名后，点击“开始搜索”按钮就会跳转到二手房信息的界面。    在“二手房信息”界面中首先可以选择想要爬取的页数，点击“开始爬取”按钮后右侧的进度条和下方的textEdit均会动态显示爬取进度    之后点击“确定”按钮会弹出“成功写入数据库”的提示框：    点击“展示”按钮后会弹出qtablewidget控件组成的界面，除了题目要求的信息外还额外爬取了“标题”、“开发商”、“房子信息”、“发布周期”等信息。    在下方的textEdit中以DataFram结构呈现数据库中的信息：    可以看到这个数据与tablewidget中的以及“北京二手房信息.csv”文件中的数据一样。  （7）日志文件：  程序运行结束后可以在项目文件夹下看到生成了logs文件夹，里面有一个2022\_11\_19.log文件，这是我写报告时运行的日志文件。  （日志文件中的内容）   1. 项目文件整体结构：      1. 代码展示：   （1）fuzzysearch()函数：      （2）数据存储到数据库中的函数insert\_table():      （3）在tablewidget中展示搜索结果函数change\_attributes(): | | | | | | |