

城市楼盘——可视化搜索引擎

一、成员介绍（第六组）

a) 各成员基本情况介绍

周依典	女	建筑与规划学院	20161150083	2016 级	城市管理
李逸歆	男	软件学院	20171120273	2017 级	信息安全
郭文彬	男	软件学院	20171120276	2017 级	信息安全
魏新	男	软件学院	20171120194	2017 级	软件工程
常泽华	男	生环学院	20171340008	2017 级	环境工程
张铭超	男	生环学院	20171340024	2017 级	环境工程

b) 具体分工

组长：

周依典 创意策划设计，负责 node 后端与 Python 爬虫，PPT 制作

组员：

魏新 设计开发讲解，负责前端页面与 Python 爬虫，代码讲解

郭文斌 开发测试讲解，负责前端可视化与 Python 爬虫，代码讲解

李逸歆 开发测试讲解，负责前端可视化与 Python 爬虫，代码讲解

张铭超 设计开发讲解，负责 node 后端与 Python 爬虫，PPT 讲解

常泽华 设计开发讲解，负责前端可视化与 Python 爬虫，PPT 讲解

二、项目作品简介

a) 选题理由

随着社会对房屋的需求增加，人们对房价原来越关注。但是由于

房价暴增时期往往高于房价下跌的时期，人们不能把握低价买房的时期。同时房价在中国也往往预示着地区的发展状况。所以原本设想，想通过创建一个平台，让人们通过查询得到这个地区各种价格的房子分布包括他的地理位置，周围交通情况，以及房价近年来的增幅情况，以便来满足人们对其的需求以及性价比的考虑

b) 项目意义

面对市场需求我们转变策划，市场上的房屋的样式，面积，朝向，规划等，这些需求与价格之间的关系也相当复杂。

所以我们在原本的基础上将目标转移到现有网站数据的爬取。我们通过爬取链家网站获取各种房屋信息。制作了一个可以通过查询城市，而得到这个城市的房价与各种需求的可视化曲线。

c) 应用场景

市场上对房屋的样式，面积，朝向，规划等需求，查询得到这个城市的房价与各种需求的可视化曲线，以此用来衡量买房的价值。

三、总体设计

1、后端使用 nodejs 来搭建，前端传来的城市参数然后使用 exec 库来调用 crawler.py 运行，crawler 接受到参数后进行爬虫作业。

2、爬虫分两步，第一步存下 1~4 页（可以增加，最好不超过 20）链家网楼盘的每个新房的地址（url），最后循环地爬取这些 url，合成这些信息，最后生成 json 文件。

3、爬虫主要采用了 BeautifulSoup 来达到易于理解的信息筛选，

即采用 css 选择器来获取内容，虽然影响了一定的速度，但比起正则表达式可维护性较高。

4、前端部分通过 ajax 请求，把城市参数传递给后端，直到后端爬虫处理完成后将数据返回后，一次对事先设定好的几个数据分析模块进行数据清洗与整理，最后将其可视化。

5、可视化部分使用了 echarts.js 库，根据不同的 option 来调整图表的表现方式。

四、特色和创新点

a) 特色

1、后端使用 nodejs 来搭建，前端传来的城市参数然后使用 exec 库来调用 crawler.py 运行，crawler 接受到参数后进行爬虫作业。

2、前端部分通过 ajax 请求，把城市参数传递给后端，直到后端爬虫处理完成后将数据返回后，一次对事先设定好的几个数据分析模块进行数据清洗与整理，最后将其可视化。

b) 创新点

1、爬虫主要采用了 BeautifulSoup 来达到易于理解的信息筛选，即采用 css 选择器来获取内容，虽然影响了一定的速度，但比起正则表达式可维护性较高。

2、可视化部分使用了 echarts.js 库，根据不同的 option 来调整图表的表现方式。

五、部分运行结果截图

1、安装 node.js，进入该文件夹的 cmd，先输入 npm install，等结束后再输入 npm start

```
Microsoft Windows [版本 10.0.17134.112]
(c) 2018 Microsoft Corporation。保留所有权利。

C:\Users\ZMC>cd C:\Users\ZMC\Desktop\lianjia

C:\Users\ZMC\Desktop\lianjia>npm install
npm WARN lianjiawang@1.0.0 No description
npm WARN lianjiawang@1.0.0 No repository field.

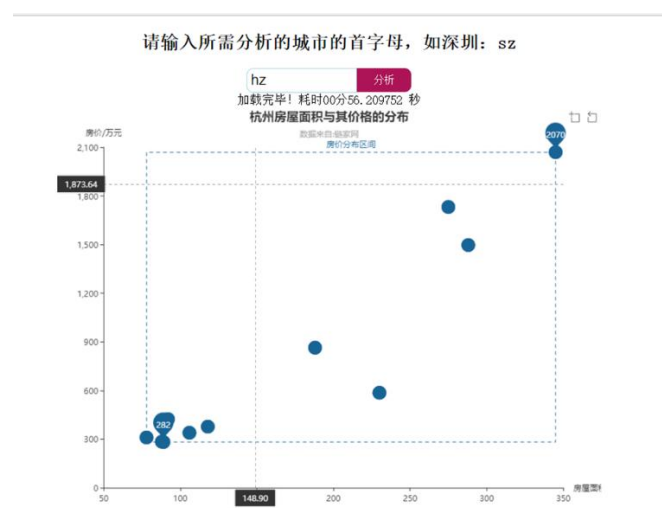
up to date in 0.298s

C:\Users\ZMC\Desktop\lianjia>npm start

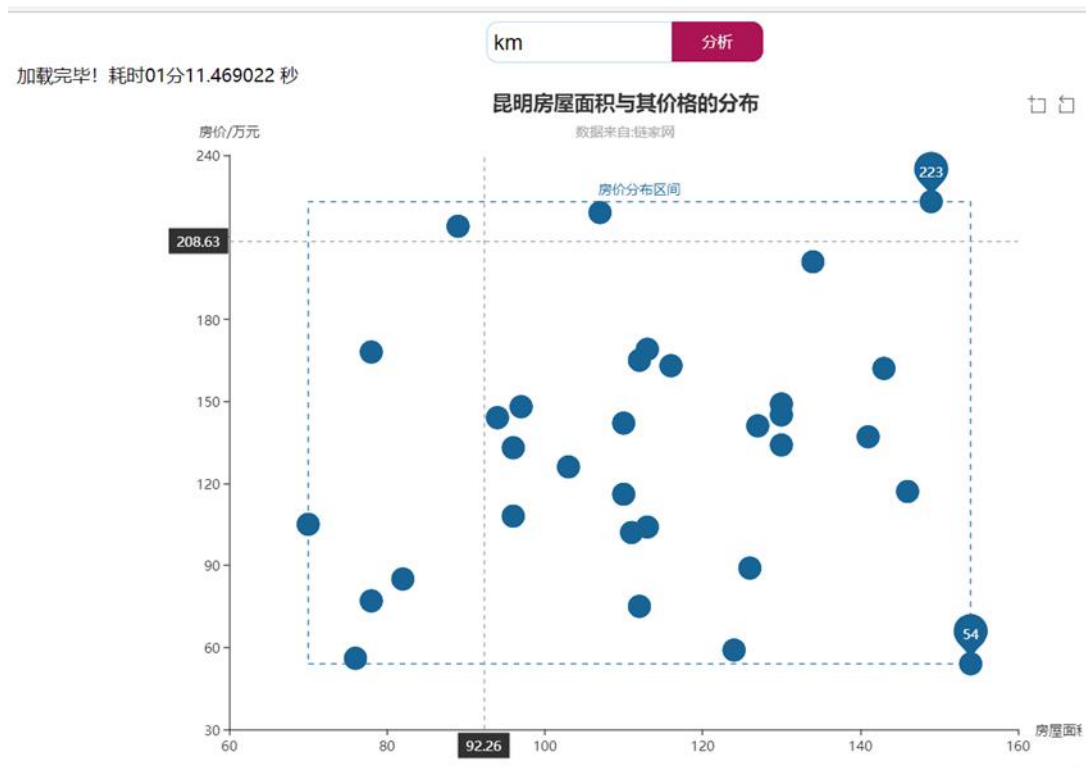
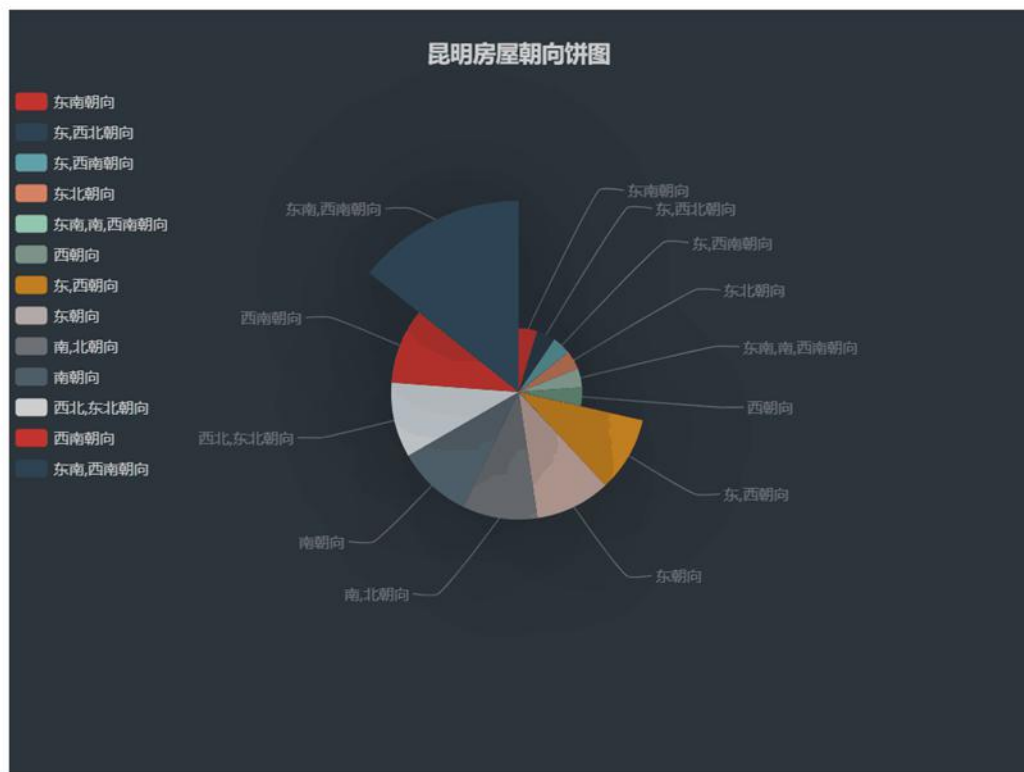
> lianjiawang@1.0.0 start C:\Users\ZMC\Desktop\lianjia
> node app.js

app listening at 3000
```

2、在浏览器（最好 chrome）输入地址 <http://localhost:3000/> 输入城市的首字母，点击分析，等待大概 1~3 分钟



3、自动出图



六、存在的问题、建议及其他需要说明的情况

原本想通过创建一个平台，让人们通过查询得到这个地区各种价格的房子分布包括他的地理位置，周围交通情况，以及房价近年来的增幅情况，以便来满足人们对其的需求以及性价比的考虑但是由于技术达不到，爬虫爬取的网站信息量不足，我们只好放弃。被其他组反映功能有所鸡肋。

后期在技术跟上的情况下能够完善平台的各种功能。

七、项目组成员的工作心得

成员一：周依典，2016 级，项目负责人

学习能力方面：有待提高，但是很努力在学习自己不擅长的方面

独立思考能力方面：有很强的的独立思考能力，能够活跃变通

发现问题能力方面：能够及时发现项目中或者团队配合中存在的问题

团队合作能力方面：组织能力强，能够配合组员工作

其它方面：创新

成员二：魏新 2017 级 代码讲解

学习能力方面：更强调自学，不懂就问

独立思考能力方面：python 是一种精简的语言，它更要求库的使用与问题的理解

发现问题能力方面：发现问题是实现项目的首要需求

团队合作方面：一个项目的完成需要通力配合

其他方面：就是需要多沟通交流，解决问题才更快更好

成员三：常泽华，2017 级 设计开发讲解

学习能力方面：自主学习了一些拓展的内容，对 Python 有了更深的理解，提高了对 Python 的应用能力。

独立思考能力方面：对于项目中存在的问题有与别人不一样的理解与处理方法，对于项目的特色与优势有自己的思考。

发现问题能力方面：在整个项目中在通过与小组成员的讨论中发现了一些问题，但是发现问题的能力还需要提高

团队合作能力方面：在整个项目的制作中每个人都有自己的任务，能将自己所负责的内容与最终的项目衔接上，在最终的汇报时分工也比较明确

其他方面：说一些自己的缺陷吧，对自己所负责的内容有较深入的了解，对于小组其他成员所负责的内容了解程度不够，这个问题需要改进。

成员四：张铭超，2017 级，设计开发讲解

学习能力方面：自学能力有提升，能够从网络和参考书籍中学习，并且将之运用至解决问题得程度。

独立思考能力方面：在实际操作中有了简单的独立思考能力，就问题进行独自地分析，联系所学，尝试解决，只是有时还需他人协助，提供思路。

发现问题能力方面：能够在简单、表象的层次提出问题，在深入根本的发现问题的源头。能力有所进步但依然有限。

团队合作能力方面：能够完成团队成员之间的沟通、讨论并产生效

果，能表达清楚自己的意思和他人所述，具体合作方面仍有所欠缺。

其他方面：真正了解了编程的具体内容，对编程的理解更加深入，编程不是一个简单的码代码的事情，在网络上对知识的不断获取，对他人成果的学习和运用，在面对问题时开阔的思路和想法都是编程里不可缺少的一部分。程序如何运行的，也应当怎样思考。