

数据抓取作业

2024-11-14

分别爬取链家官网新房和二手房的数据（各爬取从第 3 页开始的 5 页）。根据页面数据，新房要求将楼盘名称、类型、地点、房型、面积、单价和总价的数据保存到 json 文件中；二手房要求将小区名称、地点、房型等信息（字段较多可暂时不拆分）、单价和总价的数据保存到 json 文件中。

要求使用 Scrapy 框架，并在一个项目中包含两个爬虫任务。未按要求实现但也爬取到数据的酌情给分。

提交内容：编写的源程序、说明文档（不超过一页）、爬取的数据文件。

评分标准：爬取的数据文件（新房 25、二手房 25），有效的爬虫程序（25，包括反爬应对措施），使用 Scrapy 框架（10，包括是否在一个项目中包含两个爬虫任务），源程序以及说明文档（15）