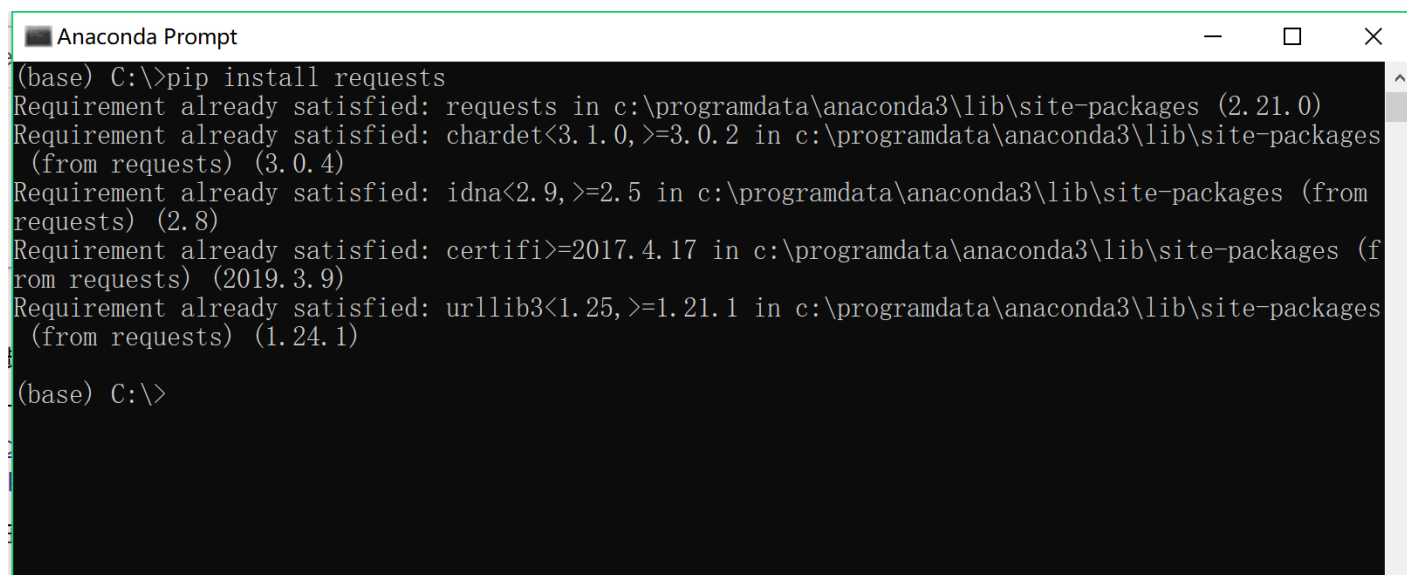


【任务】

实现这样一个程序：输入城市名，输出这个城市当日的天气情况（气温、风力等）

1. 安装 anaconda，熟悉 jupyter notebook 的用法；
2. 在 anaconda 中，安装 requests 库；



```
Anaconda Prompt
(base) C:\>pip install requests
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site-packages (2.21.0)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\programdata\anaconda3\lib\site-packages (from requests) (3.0.4)
Requirement already satisfied: idna<2.9,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests) (2.8)
Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\lib\site-packages (from requests) (2019.3.9)
Requirement already satisfied: urllib3<1.25,>=1.21.1 in c:\programdata\anaconda3\lib\site-packages (from requests) (1.24.1)
(base) C:\>
```

3. 在 jupyter notebook 中，利用 requests 库访问接口 http://wthrcdn.etouch.cn/weather_mini?city=北京，使程序实现输入一个城市名称，输出此城市当前的天气情况（气温、风力等）的功能

请输入城市，回车退出：

北京

23日星期五

高温 30℃

低温 19℃

多云

请输入城市，回车退出：

南京

23日星期五

高温 33℃

低温 24℃

多云

请输入城市，回车退出：

乌托邦

未获得

请输入城市，回车退出：

【步骤】

1. 下载安装 anaconda，下载地址：<https://www.anaconda.com/distribution/#download-section>，选择合适自己的操作系统，python 3.7 版本（网速不好的同学可以使用 <https://pan.baidu.com/s/1o79WJ2a> 下载）；
2. 在 anaconda 的终端 Anaconda Prompt 通过 `pip install requests` 安装 requests 模块；
3. anaconda 中启动 jupyter notebook（可以通过 Anaconda Prompt 输入命令 `jupyter notebook` 启动或启动 Anaconda，找到 Jupyter Notebook，点击 Launch 启动）；

4.在 jupyter notebook 中新建 python 文件，利用 requests 库的 get() 方法访问接口 http://wthrcdn.etouch.cn/weather_mini?city=北京（可提前获取用户输入，利用字符串的格式化方法，将北京替换成任意你输入的城市），获得 json 格式的天气数据，将 json 格式的天气数据使用 requests 库内置的 json() 方法转换为字典，从字典中获取城市当前的天气情况（气温、风力等），并用 print() 进行输出

【参考阅读】

Anaconda介绍、安装及使用教程: <https://zhuanlan.zhihu.com/p/32925500>

Jupyter Notebook 快速入门（上）: <https://codingpy.com/article/getting-started-with-jupyter-notebook-part-1/>

Jupyter Notebook 快速入门（下）: <https://codingpy.com/article/getting-started-with-jupyter-notebook-part-2/>

查天气文字教程: <https://python666.cn/cls/lesson/44/>

查天气视频教程: <https://mp.weixin.qq.com/s/4KXKJQFV8b7JZHPxDhzkIw>

JSON-百度百科: <https://baike.baidu.com/item/json>

【扩展阅读】

用什么写python: <https://zhuanlan.zhihu.com/p/38548187>

知乎问题 - 你为什么使用 jupyter: <https://www.zhihu.com/question/37490497>

requests 模块中文官方文档: https://2.python-requests.org/zh_CN/latest/

requests 模块介绍: https://mp.weixin.qq.com/s/97-MFV_ShERyRgTzKYjmHw

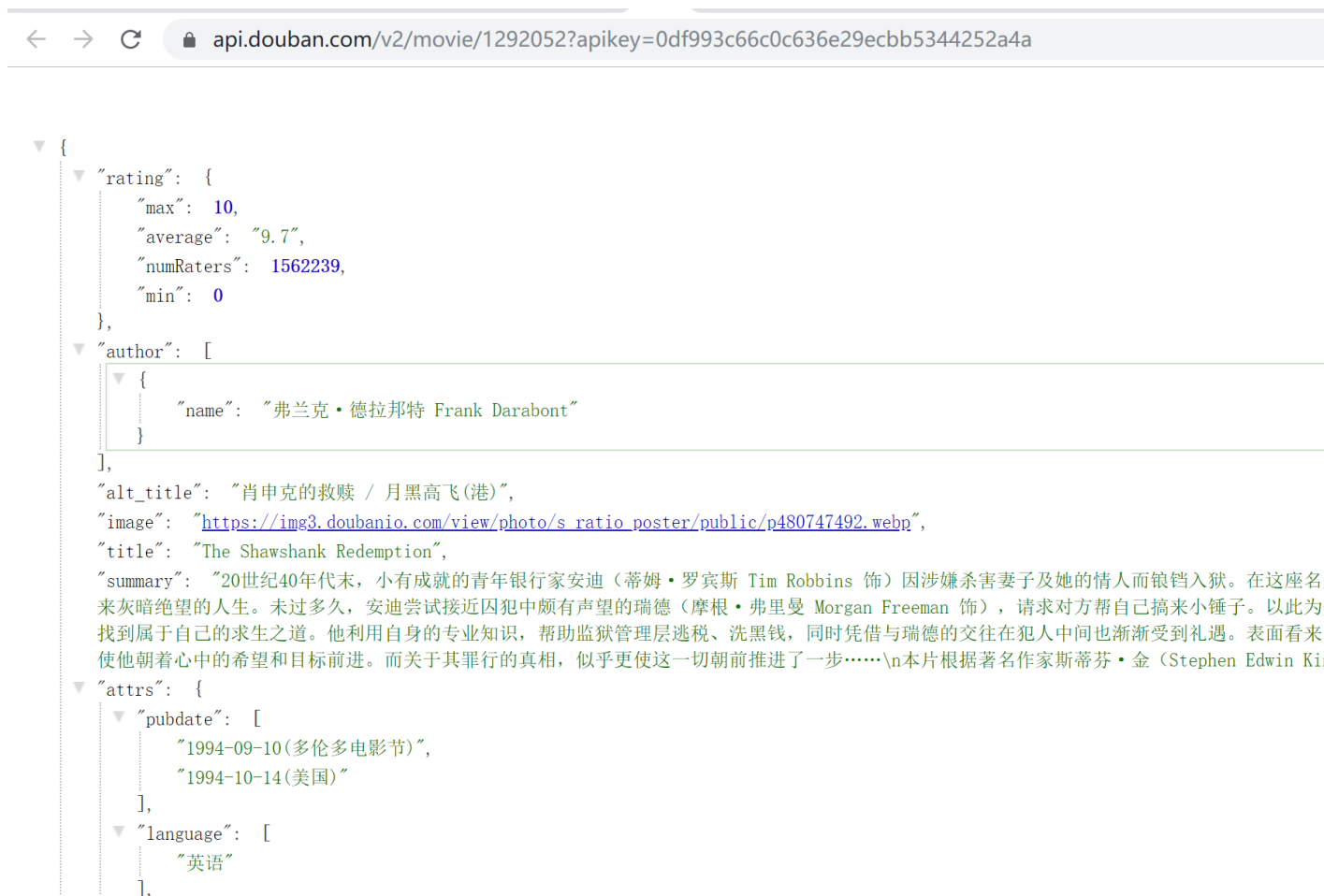
JSON在线解析: <https://www.json.cn/>

你最喜欢哪部电影？

【任务】

利用豆瓣 **api** 接口，批量抓取豆瓣电影数据并保存至本地 **csv** 文件中

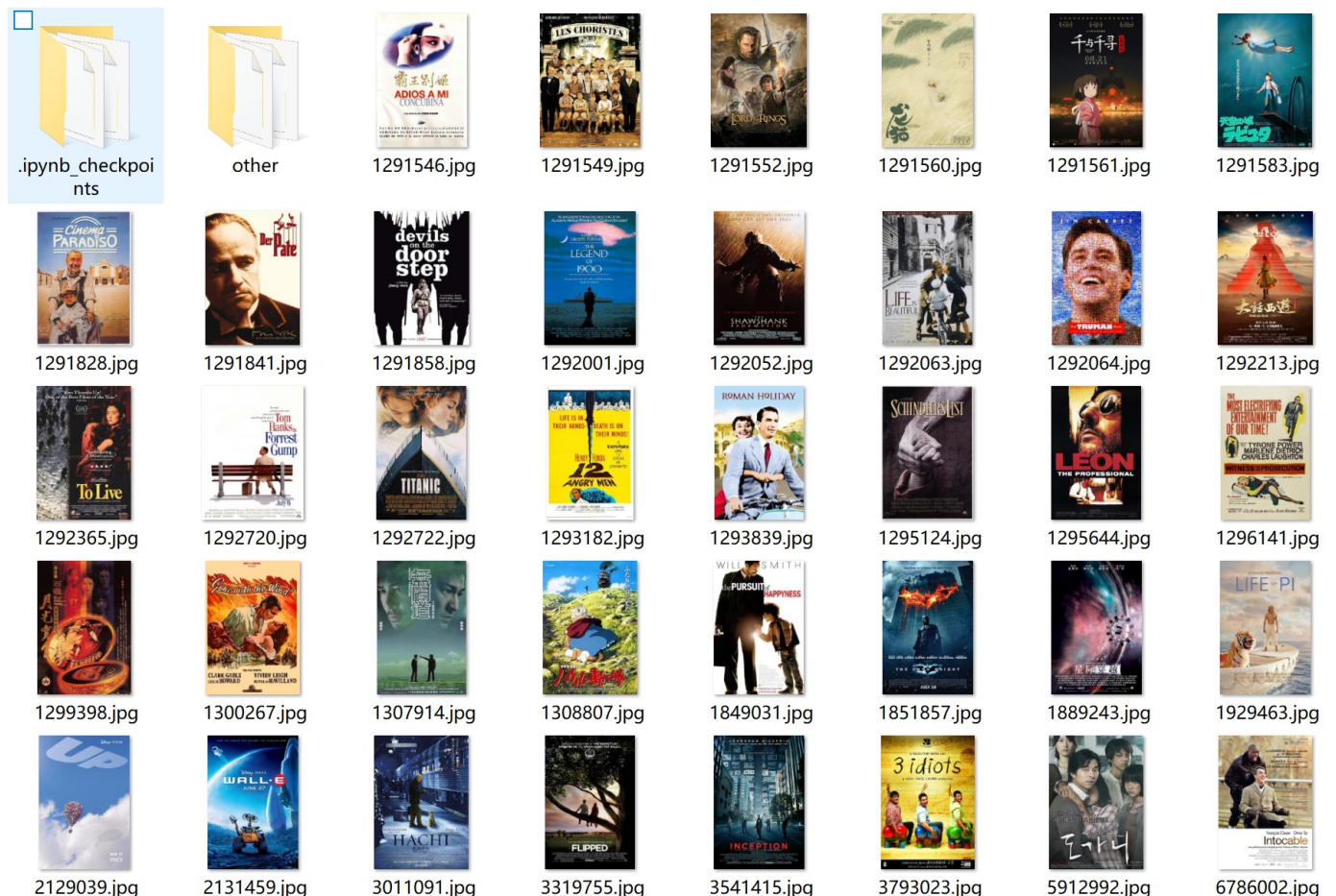
1. 利用豆瓣 **api** 接口（<https://api.douban.com/v2/movie/1292052?apikey=0df993c66c0c636e29ecbb5344252a4a>），获得电影“肖申克的救赎”的电影海报 **url** 地址，下载电影海报到本地；



2. 利用豆瓣电影 **top 250** 的 **api** 接口（<https://api.douban.com/v2/movie/top250?start=0&apikey=0df993c66c0c636e29ecbb5344252a4a>），批量获取250部电影的电影名、主演、评分等数据，保存数据到本地 **csv** 文件；

	A	B	C	D	E	F	G	H	I	J	K	L
1	1292052	肖申克的救赎	9.7	蒂姆·罗宾斯,摩根·弗里曼	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p480747492.jpg							
2												
3	1291546	霸王别姬	9.6	张国荣,张丰毅,巩俐	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p2561716440.jpg							
4												
5	1295644	这个杀手不太冷	9.4	让·雷诺,娜塔莉·波特曼,加	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p511118051.jpg							
6												
7	1292720	阿甘正传	9.5	汤姆·汉克斯,罗宾·怀特,加	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p2559011361.jpg							
8												
9	1292063	美丽人生	9.5	罗伯托·贝尼尼,尼可莱塔·	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p510861873.jpg							
10												
11	1292722	泰坦尼克号	9.4	莱昂纳多·迪卡普里奥,凯特	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p457760035.jpg							
12												
13	1291561	千与千寻	9.3	柊瑠美,入野自由,夏木真理	https://img1.doubanio.com/view/photo/s_ratio_poster/public/p2557573348.jpg							
14												
15	1295124	辛德勒的名单	9.5	连姆·尼森,本·金斯利,拉尔	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p492406163.jpg							
16												
17	3541415	盗梦空间	9.3	莱昂纳多·迪卡普里奥,约瑟	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p513344864.jpg							
18												
19	3011091	忠犬八公的故事	9.3	理查·基尔,萨拉·罗默尔,琼	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p524964016.jpg							
20												
21	2131459	机器人总动员	9.3	本·贝尔特,艾丽莎·奈特,杰	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p1461851991.jpg							
22												
23	3793023	三傻大闹宝莱坞	9.2	阿米尔·汗,卡琳娜·卡普尔	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p579729551.jpg							
24												
25	1291549	放牛班的春天	9.3	热拉尔·朱尼奥,弗朗索瓦·	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p1910824951.jpg							
26												
27	1292001	海上钢琴师	9.2	蒂姆·罗斯,普路特·泰勒·文	https://img1.doubanio.com/view/photo/s_ratio_poster/public/p511146807.jpg							
28												
29	1292064	楚门的世界	9.2	金·凯瑞,劳拉·琳妮,艾德·哈	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p479682972.jpg							
30												
31	1292213	大话西游之大圣归	9.2	周星驰,吴孟达,朱茵	https://img3.doubanio.com/view/photo/s_ratio_poster/public/p2455050536.jpg							
32												

3. (可选) 从豆瓣电影 top 250 的 api 接口中, 获取每一部电影海报的url, 批量下载电影海报到本地



【步骤】

1. 下载 Chrome 浏览器, 下载地址: <https://www.google.cn/chrome/>, Chrome 浏览器安装可在线查看 json 格式的插件“极简 Json 格式化”, 用于之后调用 api 接口后, 在线查看格式化的 json 字符串;

2. 访问豆瓣 api 接口（目前，豆瓣 api 需要 apikey，你可以使用我们提供的 apikey，也可以自己找）：
<https://api.douban.com/v2/movie/1292052?apikey=0df993c66c0c636e29ecbb5344252a4a>，利用 requests 库的 get 方法获取 json 格式的“肖申克的救赎”电影所有数据；
3. 将 json 格式的“肖申克的救赎”电影数据转换为字典，从中找到电影海报的 url 链接，再次利用 requests 访问海报的 url 链接，将二进制响应内容以二进制形式写入本地的 .jpg 文件中；
4. 利用豆瓣电影 top 250 的 api 接口（<https://api.douban.com/v2/movie/top250?start=0&apikey=0df993c66c0c636e29ecbb5344252a4a>），批量获取250部电影的数据（注意：这个接口一次只会返回 20 部电影，其中参数 start 代表返回的电影在 250 部电影中的起始位置。所以需要通过循环，给定不同的 start，来重复获取电影），循环抓取时，请使用 time 模块的 sleep() 方法控制抓取频率；
5. 对获取的豆瓣 top 250 电影数据进行处理，其中一部电影的数据生成一个列表，作为 csv 的一行保存到本地 .csv 文件；
6. （可选）从豆瓣电影 top 250 的 api 接口中，获取每一部电影海报的url，批量下载电影海报到本地，循环抓取时，请使用 time 模块的 sleep() 方法控制抓取频率，下载方法与“步骤3”一致

【参考阅读】

requests 二进制响应内容：http://cn.python-requests.org/zh_CN/latest/user/quickstart.html#id4

cookbook - 读写字节数据（图片等）：https://python3-cookbook.readthedocs.io/zh_CN/latest/c05/p04_read_write_binary_data.html

cookbook csv 读写：https://python3-cookbook.readthedocs.io/zh_CN/latest/c06/p01_read_write_csv_data.html

Python time sleep()方法：<https://www.runoob.com/python/att-time-sleep.html>

【扩展阅读】

知乎问题 - 什么是 api 接口：<https://www.zhihu.com/question/38594466>

csv 中文官方文档：<https://docs.python.org/zh-cn/3/library/csv.html>

开始进一步的抓取

【任务】

在智联招聘中，抓取 **python** 相关职位信息和详细职位描述，保存数据到本地 **csv** 文件

1. 在智联招聘（<https://www.zhaopin.com/>）中，抓取 **python** 相关职位一定数量的招聘信息；

```
    "typeSearch": 0,
  "results": [
    {
      "number": "CC273670418J00207713211",
      "jobName": "python开发工程师",
      "company": {
        "name": "北京捷科智诚科技有限公司",
        "number": "CZ273670410",
        "type": {
          "name": "民营"
        },
        "size": {
          "name": "1000-9999人"
        },
        "url": "https://company.zhaopin.com/CZ273670410.htm"
      },
      "city": {
        "items": [
          {
            "name": "上海",
            "code": "538"
          },
          {
            "name": "浦东新区"
          }
        ],
        "display": "上海-浦东新区"
      }
    },
  ],
}
```

2. 进入每一个招聘详情页，抓取该职位的详细职位描述。最后将所有结果以 **csv** 形式保存在本地

1	15-20K: 高级程序员/软件开发/.net/C# (港资公司), 15K-20K, 本科, 3-5年, https://jobs.zhaopin.com/CC133675644J00407323505.htm , 全职, 上海中原物业顾问有限公司, 外商独资, 10000人以上, "1、年龄: 28-36岁; 2、专业: 本科以上学历, 计算机或相关专业; 3、知识技能: 3.1、精通C#, 熟练使用VS.NET 2003、VS.NET2008、VS.NET 2012、3.2、熟练使用主流数据库sqlserver 2000、sqlserver2008、mysql、oracle 3.2、熟练使用主流数据库sqlserver 2000、sqlserver2008、mysql、oracle 3.3、编写高效的sql脚本、视图、存储过程 3.4、熟练使用div、javascript、css、html3.5、熟练使用常用代码管理工具sourcesafe、svn3.6、熟悉MVC架构 3.7、熟悉项目管理的基本流程: 需求、开发、测试、代码review、风险控制等环节 3.8、了解工作流的基本原理、使用过K2工作软件进行开发 3.9、了解人事管理软件的业务知识4、工作经验: 4.1、3年以上大型项目开发经验5、素质要求: 5.1、具备良好的沟通协调能力和组织能力、表达能力、较强的执行力; 5.2、良好的团队精神, 工作积极主动、责任心强; 5.3、能够有吃苦耐劳的精神 (一) 主要工作职责: 1、负责客户需求的沟通、开发2、负责数据库的性能优化和代码优化 联系方式: 应聘邮箱: ***** "
2	
3	机器人轨迹规划算法工程师, 20K-30K, 硕士, 3-5年, https://jobs.zhaopin.com/CC822216530J00301929106.htm , 全职, 佳奕彼安 (上海) 机器人科技有限公司, 民营, 20人以下, 1、研究生及以上学历或3年以上工作经验, 自动化、机械工程、计算机、软件工程等相关专业; 2、丰富的编程经验, 熟悉C/C++, 熟悉Matlab、Visual Studio等。 3、熟悉机器人的运动学及轨迹规划算法等 4、熟悉工业机器人基本操作; 5、有机器人开发相关工程经验者优先;
4	
5	软件研发工程师, 13K-18K, 本科, 3-5年, https://jobs.zhaopin.com/CC156501624J00422274001.htm , 全职, 上海唐盛信息科技有限公司, 民营, 100-499人, "岗位职责1. 负责需求分析、程序研发、模块调试、集成和维护工作; 2. 负责解决软件开发过程中的技术问题; 3. 根据规范要求编写相关的实施文档、使用手册、用户培训文档等; 职位要求1. 专科及以上学历, 3年以上开发经验, 计算机相关专业毕业, 有通信行业工作背景者最优; 2. 熟悉Linux操作系统, 能够执行基本的操作命令和编译程序; 3. 熟练使用C/C++/Java/python编程, 若仅掌握其中一种语言, 需要达到高级水平; 4. 熟练使用Redis、Mango、MySQL等数据库; 5. 熟悉Docker开发和部署环境; 6. 较强的英语读写能力; "
6	
7	数据分析及开发工程师 (Python), 10K-16K, 本科, 1-3年, https://jobs.zhaopin.com/CC000584094J00315376302.htm , 全职, 上海喜士多便利连锁有限公司, 民营, 1000-9999人, 工作描述: 1、根据业务单位需求完成数据收集和整理; 2、配合大数据分析师完成数据处理算法和应用的开发; 3、研究当前成熟的AI技术对公司业务改善的可行性; 岗位要求: 1、计算机相关专业, 精通python编程; 2、对自然语言处理有兴趣; 3、可以设计编写爬虫软件收集互联网数据, 最好有一定的大数据处理经验。4、熟悉关系型数据库 (Oracle/MySQL) 和SQL语法; 5、有良好的代码编写和管理习惯, 熟悉一种主流的版本控制系统 (SVN/Git); 6、善与人沟通、责任心强、具有团队合作精神、具备良好的自学能力和独立思考解决问题的能力。联系方式: 杨小姐 000000000000
8	
9	python开发工程师, 8K-10K, 大专, 1-3年, https://jobs.zhaopin.com/CC839265400J00260949205.htm , 全职, 上海蜗兴科技有限公司, 民营, 20-99人, 任职要求: □1、计算机相关专业大专及以上学历; 2、有Python开发经验, 熟悉Django、SQLAlchemy等框架; 3、熟悉SQL能进行简单的数据统计、分析; 4、熟悉Web前端技能, 包括HTML/CSS/JavaScript; 5、对Web后端技术架构有基本的理解; 6、有良好的团队合作精神, 有责任感, 有较好的沟通能力; 7、喜欢尝试新的技术, 对技术趋势和思路能提出自己的见解。岗位职责: 1、参与公司网站开发、维护、实施、测试; 2、负责现有网站的维护和升级; 3、负责解决开发过程中的技术问题; 4、根据开发规范与流程独立完成编码、测试及相关文档; 福利待遇: 1、丰富多彩的集体活动 (员工旅游, 节日聚会, 生日趴, 年会等等。) 2、入职购买五险一金、周末双休、员工旅游、员工体检、带薪年假。
10	
11	Python/C++/C# 软件开发工程师, 8K-10K, 本科, 不限, https://jobs.zhaopin.com/CZ729442480J00305178301.htm , 全职, 上海金柳投资管理有限公司, 民营, 20人以下, 岗位职责: 1、参与自动交易平台的设计、开发和测试2、参与高性能数据存储解决方案的设计、开发和测

【步骤】

1. 在智联招聘 (<https://www.zhaopin.com/>) 中, 输入职位关键词: **python**, 抓取这个职位一定数量的招聘信息 (职位为异步加载, 需要从开发者工具中找到 json 格式的数据来源), 抓取的信息可以包括薪水、学历要求、工作经验要求、招聘详细url、是否全职、公司名、公司类型、公司规模等;
2. 在完成步骤1, 抓取招聘列表的基础上, 进入每一个招聘详情页, 抓取该职位的详细职位描述, 注意抓取时要控制频率;
3. 一个职位信息作为一行, 将所有结果以 csv 形式保存在本地

【参考阅读】

HTML <a> 标签: https://www.w3school.com.cn/tags/tag_a.asp

换个方式刷微博

【任务】

抓取 微博热点 页面并保存为本地的 **HTML** 页面


1. 学习使用 **Chrome** 开发者工具，找到微博热点（<https://weibo.com/a/hot/realtime>）在请求过程中所用的 **cookie** 和 **User-Agent**；

The screenshot displays the Weibo website's trending topics page. The browser's address bar shows the URL `weibo.com/a/hot/realtime`. The page lists several trending topics, including `#台风白鹿#`, `#有匪#`, `#小欢喜细节#`, `#磊儿越来越开朗#`, and `RNG Liquid`. On the right side, the Chrome DevTools Network tab is open, showing a list of requests. The request `realtime` is selected, and its headers and cookies are visible. The cookies include `SUBP=0033WrSXqPxfM72-WS9JqgMF55529F8`, `SUB=2AkMru108f8`, `SINAGL`, `_s_tent`, `V=1566528098884:15:5:1:8818912436357.828.15`, `Ugrow-G0=5c7144e5`, `YF-V5-G0=7e17aef9f7`, `UOR=,bbs.51testing.co`, and `WBStorage=f54cf4e4362237da|undefined`. The host is `weibo.com` and the `If-Modified-Since` is `Fri, 23 Aug 2019 09:29:35 GMT`.


2. 将 **cookie** 和 **User-Agent** 参数加入 **requests** 的 **headers** 中，利用 **requests** 抓取 微博实时热点 并将返回内容保存为一个本地的 **HTML** 页面



#台风白鹿#


 [中央气象台](#) 8月21日 16:04 *~1224* ☐ 3568 ☐ 56900

#有匪#


 [娱扒爷](#) 8月18日 21:35 *~273* ☐ 682 ☐ 61993



#小欢喜细节#

 [影视圈老顽童](#) 8月22日 22:13 *~284* ☐ 1159 ☐ 21465

#磊儿越来越开朗#

 [热剧追踪控](#) 今天 07:54 *~1553* ☐ 2474 ☐ 180081

【步骤】

1. 用 Chrome 浏览器打开微博热点（<https://weibo.com/a/hot/realtime>），在网页上右键，选择检查，打开开发者工具；
2. 打开后，选择 Elements 标签页，找到 重新刷新网页，找到请求页面（<https://weibo.com/a/hot/realtime>）的 Request Headers 信息，复制 cookie 和 User-Agent 用于之后的请求；
3. 将复制的 cookie 和 User-Agent 参数以字典形式加入请求的 headers 中（添加 headers 方法见参考阅读：[requests 官方文档 - 定制请求头](#)），用 requests 库请求微博热点；
4. 请求完成后，将返回结果写入到一个 .html 文件，试试在浏览器打开刚刚保存的文件

【参考阅读】

chrome 开发者工具介绍：<https://zhuankun.zhihu.com/p/39340856>

requests 官方文档 - 定制请求头：http://cn.python-requests.org/zh_CN/latest/user/quickstart.html#id6

【扩展阅读】

什么是 HTTP Headers? - GuYoung: <https://kb.cnblogs.com/page/119118/>

HTTP响应头和请求头信息对照表：http://tools.jb51.net/table/http_header

HTML 教程：<https://www.runoob.com/html/html-tutorial.html>

登录后玩玩知乎

【任务】

登录知乎后抓取你的知乎时间线内容并将结果保存在本地的 csv 文件中

1. 打开并登录知乎（<https://www.zhihu.com>），在 chrome 开发者工具中，找到知乎时间线中异步数据来源的 url 和对应登录后的 cookie；

知乎

首页 发现 等你来答

有哪些适合学生使用的 App

提问

推荐 关注 热榜

有哪些物理学上的事实，没有一定物理学知识的人不会相信？

1000米恐惧症：非牛顿流体，你越用力它越硬，你不用力它就变成软塌塌的一坨。注意卫生的话玩完了还能吃，做家庭小实验必备。制作非牛顿流体也很简单，淀粉+水就行，注意不是和面用的小麦面粉，水尽量少加就可... 阅读全文

如何评价杨和苏获得新说唱冠军？

臧默：站客观角度评价一下，首先新秀就不谈了，第一个淘汰都是意料之中，battle mc出身走到这一步已经很厉害了，作品性还是不够成熟，弹药... 阅读全文

被女舔狗舔是一种怎样的体验？

我不知道你是谁：前女友第一次见我，是在学校食堂，匆匆一瞥之后和室友放下饭盘立刻追我而去，奈何追至楼下我踏上自行车一骑绝尘.....她和室... 阅读全文

码云Gitee

企业为什么要统一精细化代码管理，管理者应该知道这些真相

代码管理不仅仅是代码托管，而是在于代码管理的同时与项目管

写回答

我的稍后读

我的草稿

Live

专栏

我的收藏

我关注的

我的邀请

我的余额

站务中心

版权服务

刘看山·知乎指南

应用·工具

Network

XHR

recommend?session_token=

accept-encoding: gzip, deflate, br

accept-language: zh-CN,zh;q=0.9,en;q=0.8

cookie: _zap=088aaa30-35c4-4de1-89eb-3b18b2c0="AKD1816vQA-PtNbuNu8PeYyXgoJ5A4sk96w=|14"; _xsrf=gy5NQbmbP3NdZ1cx7WMyWn1vbqUm1P2U1:0|10:1556116755|4:z_c0|92:Mi4xeEJNcEFBQUeVhOUFeEVLBQUFCZ0FsVk5FOE90WFFEVwx3TjBFQwN3RmZlEzQ3JrNmF3|7a06a0d1e79ce254c0dd9ba9a9a8c37891bd2cf6f5584ceb48b7b7"; __gads=ID=2b9364:T=1558351421:S=ALNI_MZvhdv6MyTmPTexj6PA; __utmv=51854390.100-1|2=registration1217=1^3=entry_date=20131217=1; __utma=51894488.1558436044.1558436044.1559127908.2; 54390.1559127908.2.2.utmcsr=zhihu.com|utmc=al)|utmcmd=referral|utmctt=/people/fu-lan-

2. 将 User-Agent 和具有登录信息的 cookie 加入 requests 的 headers 中，用 requests 库请求知乎时间线异步的数据来源，请求完成后，将请求数据保存为一个 .csv 文件

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	6_1566552	电影	7.92E+08	12444	732	599	224068	1.57E+09	47159584	为什么熊猫能跟功夫扯上关系	80	216	3	1.47E+09
2														
3	7_1566552	星巴克	7.85E+08	86	43	12	12714	1.57E+09	22035105	在星巴克工作是怎样一种体验	428	7327	12	1.38E+09
4														
5	8_1566552	照片	7.75E+08	489	260	63	254185	1.56E+09	2.68E+08	你最照骗的一张照片是	19982	50479	61	1.52E+09
6														
7	9_1566552	博士	7.5E+08	2093	495	198	359606	1.56E+09	3.34E+08	为什么读博士这么惨还	293	1378	14	1.56E+09
8														
9	10_156655	技术	7.46E+08	55887	1229	10837	2036168	1.56E+09	3.34E+08	技术一旦被用来作恶，	962	26676	62	1.56E+09
10														
11	11_156655	丰田 (Toy	7.95E+08	32	36	6	15615	1.57E+09	3.37E+08	丰田，本田，日产谁的	9	15	0	1.56E+09
12														

【步骤】

- 用 Chrome 浏览器打开知乎（<https://www.zhihu.com>）并登录。登录后进入 Chrome 开发者工具，选择 Network 标签页的XHR 标签，下拉加载网页页面后，找到异步请求的数据来源（https://www.zhihu.com/api/v4/search_v3?...）；
- 将数据来源中，记录了网站登录态信息的 cookie 和浏览器信息 User-Agent 复制下来用于之后的请求；
- 用 requests 库请求知乎时间线异步的数据来源，将 cookie 和 User-Agent 加入请求的 headers 中；
- 请求完成后，将 json 格式的返回数据转换为字典，每一项数据生成一个列表，作为一行，写入进 .csv 文件

【参考阅读】

chrome浏览器fiter中的XHR作用是什么：https://blog.csdn.net/qq_26296197/article/details/79616480

HTTP 方法：GET 对比 POST：https://www.w3school.com.cn/tags/html_ref_httpmethods.asp

【扩展阅读】

菜鸟教程 - AJAX: <https://www.runoob.com/ajax/ajax-tutorial.html>

聊一聊 cookie - ruoyiqing: <https://segmentfault.com/a/1190000004556040>

Javascript异步编程的4种方法: <http://www.ruanyifeng.com/blog/2012/12/asynchronous%EF%BC%BFjavascript.html>

从页面能得到什么

【任务】

从上海链家的页面中批量抓取租房房源信息，保存在本地 csv 文件中

1. 查看 上海链家（<https://sh.lianjia.com/zufang/>）的网页结构，找到房源信息所在位置，为 bs4 解析网页做准备；

The screenshot displays the Shanghai Lianjia rental website interface. The main content area shows three rental listings, each with a thumbnail image, title, location, area, and price. The prices are highlighted in red: 430, 360, and 480. The right side of the image shows the browser's developer tools, specifically the 'Elements' panel, which displays the HTML structure of the page. The HTML structure shows the following elements:

- `<div class="content_list-item" data-el="listItem" data-house_code="SH2328153178211885056" data-brand_code="200306004772" data-ad_code data-c_type="1" data-position="0" data-total="19611" data-fb_expo_id="217324064515616769" data-t="default" data-strategy_id data-click_position="0" data-ad_type="0" data-distribution_type="203500000001">`
- `<div class="content_list-item" data-el="listItem" data-house_code="SH2326134040845369344" data-brand_code="200306004772" data-ad_code data-c_type="1" data-position="1" data-total="19611" data-fb_expo_id="217324064515616770" data-t="default" data-strategy_id data-click_position="1" data-ad_type="0" data-distribution_type="203500000001">`
- `<div class="content_list-item" data-el="listItem" data-house_code="SH2326134040845369344" data-brand_code="200306004772" data-ad_code data-c_type="1" data-position="1" data-total="19611" data-fb_expo_id="217324064515616770" data-t="default" data-strategy_id data-click_position="1" data-ad_type="0" data-distribution_type="203500000001">`

The HTML structure also includes the following elements:

- ``
- `<div class="content_list-item-main">`
- `<p class="content_list-item-title twoline">`
- `<p class="content_list-item-des">`
- `<p class="content_list-item-brand online">`
- `<p class="content_list-item-time online">`
- `<p class="content_list-item-bottom online">`
- ``

2. 利用 bs4 解析包含房源信息的网页，得到每一个房源的详细信息，将抓取的信息保存至csv

	A	B	C	D	E	F	G	H	I	
1	整租·仙逸小区 1室1厅 南	/zufang/SH2328153178211	长宁-仙霞	41m²	南	1室1厅1卫	高楼层	近地铁,新上,随时看房		
2										
3	整租·上中路458弄 1室1厅 南	/zufang/SH2326134040845	徐汇-植物	38m²	南	1室1厅1卫	高楼层	新上,随时看房		
4										
5	整租·和平小区(徐汇) 2室1厅 南	/zufang/SH2323737071614	徐汇-华东	75m²	南	2室1厅1卫	低楼层	随时看房		
6										
7	整租·花苑村丁香园 1室1厅 南	/zufang/SH2316386598926	徐汇-康健	47m²	南	1室1厅1卫	高楼层	近地铁,精装,随时看房		
8										
9	整租·长桥三村 2室1厅 南	/zufang/SH2307750683245	徐汇-长桥	58m²	南	2室1厅1卫	中楼层	随时看房		
10										
11	整租·淡水路442弄 2室1厅 南	/zufang/SH2310669801573	黄浦-新天	46m²	南	2室1厅1卫	中楼层	近地铁,随时看房		
12										
13	整租·汇成二村 1室0厅 南	/zufang/SH2317354078369	徐汇-上海	33m²	南	1室0厅1卫	高楼层	精装,随时看房		
14										
15	整租·华建一街坊 2室1厅 南	/zufang/SH2306213071539	徐汇-华泾	56m²	南	2室1厅1卫	高楼层	随时看房		
16										
17	整租·南丹小区 1室1厅 东南/南	/zufang/SH2298455869585	徐汇-徐家	37m²	东南 南	1室1厅1卫	低楼层	近地铁,随时看房		
18										
19	整租·汇成五村 2室1厅 南/北	/zufang/SH2308484721056	徐汇-上海	59m²	南 北	2室1厅1卫	高楼层	随时看房		
20										
21	整租·日晖二村 1室1厅 南	/zufang/SH2318110068161	徐汇-斜土	38m²	南	1室1厅1卫	低楼层	近地铁,随时看房		
22										
23	整租·田林十二村 1室1厅 南	/zufang/SH2298569163675	徐汇-田林	30m²	南	1室1厅1卫	中楼层	近地铁,随时看房		
24										
25	整租·田林十四村 1室1厅 南	/zufang/SH2298574048889	徐汇-田林	37m²	南	1室1厅1卫	高楼层	近地铁,随时看房		
26										
27	整租·长桥一村 1室0厅 东南	/zufang/SH2314210525871	徐汇-长桥	42m²	东南	1室0厅1卫	高楼层	精装,随时看房		
28										
29	整租·花苑茶花园 1室1厅 南	/zufang/SH2311635846958	徐汇-康健	33m²	南	1室1厅1卫	中楼层	精装,随时看房		
30										
31	整租·百龙小区 2室1厅 南	/zufang/SH2326134993455	徐汇-植物	85m²	南	2室1厅1卫	高楼层	新上,随时看房		

【步骤】

1. 利用 requests 库抓取上海链家（<https://sh.lianjia.com/zufang/>）一定页数的房源信息；
2. 打开 chrome 的开发者工具，选择 Elements 标签页查看 上海链家 的网页结构，找到房源信息所在的网页元素位置，为 bs4 解析网页做准备；
3. 利用 bs4（<https://www.crummy.com/software/BeautifulSoup/bs4/doc/index.zh.html>）解析网页，用 find_all() 方法找到所有所有房源共有的 class，对所有房源进行遍历，依次寻找每个房源的详细页链接、位置、面积、朝向和户型等信息；
4. 将每个房源所有抓取的信息作为一行，保存至 .csv 文件

【参考阅读】

浏览器如何查看网页源代码：<https://jingyan.baidu.com/article/09ea3ede21b18cc0aede39d7.html>

Beautiful中文官方文档：<https://www.crummy.com/software/BeautifulSoup/bs4/doc/index.zh.html>

【扩展阅读】

HTML 教程：<https://www.runoob.com/html/html-tutorial.html>