

作业 1: 爬虫

姓名: 白植权 学号: 221050020

1. 作业简介

本作业完成的是一个微博数据爬虫，它的主要功能是爬取微博上关于特定话题（如“研究生”“杭州”“周杰伦”）的微博内容。代码会创建一个 CSV 文件，用于存储爬取的数据，包括用户名和对应的微博内容。代码在预设的最大迭代次数（这里设为 100 次）内，构建并访问微博的搜索页 URL，这个 URL 根据特定话题进行构建。每次访问页面时，爬虫会获取该页的微博数据，将其中的用户名和评论内容提取出来，并保存到 CSV 文件中。

2. 爬虫 Python 部分代码:

```
# encoding=gbk
import requests
import pandas as pd
import time
import re
from tqdm import trange
topic = "大学生"
epochs = 100
def get_data():
    data2 = pd.DataFrame([{"id": "comment"}])
    data2.to_csv('%s.csv' % topic, header=False, index=False) # 导出数据到{topic}.csv
    #. . . 获取网页header 和 cookie
    .
    .
    .
    print("开始爬取数据")
    # 爬取数据
    for ii in trange(epochs):
        url_base = . . . # url 太长,在此省略
        url = url_base + str(ii + 1)

        json = requests.get(url, headers=header, cookies=Cookie).json()
        try:
            if len(json['data']['cards']) == 0:
                print("爬取结束")
                return
            for jj in range(len(json['data']['cards'])):
                text = re.sub(r'<.*?>', '', json['data']['cards'][jj]['mblog']['text'])
                # 去英文数字和部分符号
                data1 = [(json['data']['cards'][jj]['mblog']['user']['screen_name'], text)]
                data2 = pd.concat([data2, pd.DataFrame(data1)], ignore_index=True)
        except Exception as e:
            print(f"Error: {e}")
    data2.to_csv('%s.csv' % topic, header=False, index=False)
```

```

data2 = pd.DataFrame(data1)

data2.to_csv('%.csv' % topic, header=False, index=False, mode='a+')

# 导出数据到{topic}.csv

except:

    print("第%d页抓取失败" % ii)

time.sleep(3)

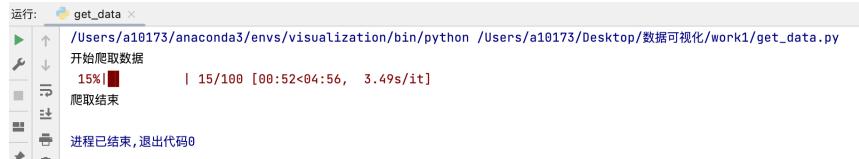
# 防止高频连续爬取被禁止

if __name__ == "__main__":
    get_data()

```

3. 运行结果：

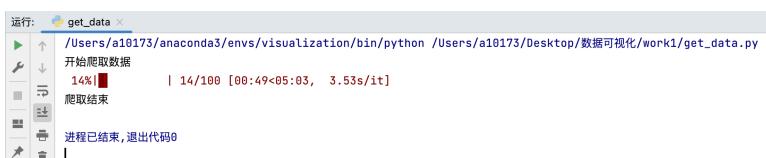
(1) 主题为“大学生”时：



爬取资料部分截图如下：(此为 2023 年 5 月 25 日爬取数据)：

大学生	
id	comment
江西卫视	高校女生跳舞，上演《恋爱告急》，又美又飒甜度超标，网友：果然内娱还得看大学生。#恋爱告急##跳舞##大学生# 江西卫视的微博视频
演员刘妍...	一晃月中#大学生# 的排练日常 🎉
梅子的酱	燃烧吧！我滚烫的青春！！#运动会##大学生# 北京·中国传媒大学
猩猩新闻	【#女生穿三次元连衣裙跑接力赛#】4月8日，安徽池州，女生穿三次元连衣裙跑接力赛，整个观众席都沸腾了。据当事人室友发，裙子是我借给她的，反正要毕业了就玩点刺激的。#大学生# 大学生 猩猩新闻的微博视频
考研的元气落落	论文实用工具希望大家不要被毕业论文折磨#毕业论文##大学生#
茉莉花茶有如玉的姐	#大学生 #7万没想到#情侣 #大学生情侣根据的生活，破防了 哎，女孩子好可怜 茉莉花茶有如玉的甜的微博视频
我是李T33	有人在看段子，有人在照镜子#大学生##普通话##微博计划# 我是李飞33的微博视频
极品天菜	烧饭我们经常煮了#网红##大学生 #颜值男孩#精神状态良好 极品天菜的微博视频
升值计	天涯论坛无法访问了，对于很多70后、80后来说，天涯曾经是他们的精神家园，我想就此来谈一谈互联网时代的财富密码，这个财富密码，我把它称之为，互联网学生三定律，简单来说，就是四个字，大学生。为了说明这个
轻阅朴读	#转发赠书##好书推荐# 横扫各大#推理#榜一度超越东野圭吾《假面舞会》/伊坂幸太郎# #小岛秀夫#力荐→浅仓秋成引爆狂潮现象级IP# 《六个说谎的大学生》讲述6个#校招##大学生#为竞争名企offer上演的一场沉没
三联中读	【#中读 每天一本书#转发赠书】今天送的是#鄙视文化#出版的现象级推理小说《六个说谎的大学生》横扫各大推理榜单一度超越东野圭吾《假面舞会》/伊坂幸太郎# #小岛秀夫#力荐#名酒连载#舞台剧火热上演#电影街
江上的路	#大学生##大学生#e9##高校##985##211# 这张图做的好，了解一下我国大学的层次：首先是C9，次之是985，再次之是部属211，再次之是省属211，其中985与211差别巨大。特别是省属211，因为当时为了照顾和平衡地区差异
墨脱财经	#大学生 #某西部985大学生生活消费调查# 数字背后的真相是残酷而真实。 9%的大学生消费在1000以下，可以肯定这些孩子家庭基本是贫困的，每天30块就是温饱水平，别的消费基本是谈不上的。 1000到2000的
郑房有问有答	一个朋友在郑州经营快捷酒店的，上个月说，他的前台拍了一个空乘专业的漂亮的小姐姐，那喜悦之情溢于言表。嘴巴都裂到耳朵上了。的确，长得漂亮，大长腿，穿着制服往前台一站，感觉自己酒店的档次都提高了。当时，
HOMO猫	原神不是，我很不能理解，我一没错犯，二没有挂科，而我们辅导员因为我在空间里面经常发原神，所以对我不满，想要把我换掉？是我们班团支书，智慧团建也是了，换团长书很麻烦的啊？#原神##大学生#
青蛙王子波仔	优缺点完美集一身#表情包#今日memes##大学生#
大学笔记墙	大学生一定要会打破信息差啊！少走弯路太香了！#大学生#
Just_Jerry_	毕业典礼现场的我 🎉 优秀毕业生提名环节突然接到名字时的surprise到现在脑子都还是嗡嗡的❶ps真的很喜欢伊斯曼大礼堂的庄重感！❤#大学生##毕业典礼##毕业季##2023毕业季# 美国 Kodak Hall at Eastman Theatre
是安安安a	女大学生被人包围了一天发生了灾祸！会动的我 #大学生##今年的五一疯了吗#北京·王府井步行街
七七爱生气	吃饭睡觉写论文#大学生#
周一和周二_twins	No way 今日的男孩来自校园一周一和周二 @周一周二_yy 我想看看你的春天#大学生##令人心动的日常# 周一和周二_twins的微博视频
比空白更空白的洁白	妈妈故意作诗忘记生活费，男生准备回学校 跟妈妈告别，徘徊犹豫很久 不知道怎么开口要钱，网友：每次问家里要钱总有一种负罪感😊 (来源：@啊干先生 (上学版)) #社会# 母爱 #大学生 #内容过于真实 #真实事件 #你们
京Gene	来看新学长的昨天穿搭 #adonijing2020 #20年感谢 #网球服#体育生 #介绍一个小哥哥 ##大学生 #
湖北专升本网	我就知道你一定上岸。朋友圈文案都给你准备好了#湖北专升本# #大学生#
今日最hot	近日重庆，被小姐妹睡前准备的恶搞操作逗笑了 被小姐妹种草，这蚊帐也太可爱了#大学生# 今日最hot的微博视频
赛茜Cecilia	如果你也有「鸿鹄之志」，就来加入我们吧！ 今年是我成为CYOT战略部联席主管的第4年，我们CYOT鸿鹄计划也进入4.0了。很开心在这儿分享项目的最新动态与招新信息😊 CYOT鸿鹄计划，是中国首家赋能第一代大学生
微博校园	谢谢你，简称侠#简历##大学生#
Miss周星星	3个穷坑学生的穷坑刺客！坑货！坑货！第四期来了已经要到第四期，这个系列比想象中累每次发出去，需要的还没看到，已经来了希望国货们都别着急，产品好坏消费者心里都有数咱们国货缺的不是情怀而是真实好用的
香槟玫瑰173	日常我要山 要海 要自由开玩笑的我要上大学#发展##日常#
noya姐	#四川文理学院吃相太难看# 各位新生学弟学妹们 快跑！坑货！第四期来了已经要到第四期，这个系列比想象中累每次发出去，需要的还没看到，已经来了希望国货们都别着急，产品好坏消费者心里都有数咱们国货缺的不是情怀而是真实好用的
Miss周星星	3个穷坑学生的穷坑刺客！坑货！坑货！第四期来了已经要到第四期，这个系列比想象中累每次发出去，需要的还没看到，已经来了希望国货们都别着急，产品好坏消费者心里都有数咱们国货缺的不是情怀而是真实好用的
香槟玫瑰173	日常我要山 要海 要自由开玩笑的我要上大学#发展##日常#
noya姐	#四川文理学院吃相太难看# 各位新生学弟学妹们 快跑！坑货！第四期来了已经要到第四期，这个系列比想象中累每次发出去，需要的还没看到，已经来了希望国货们都别着急，产品好坏消费者心里都有数咱们国货缺的不是情怀而是真实好用的
良田的田	路过的每一个人至少说自己的优点：_____ "#自信##求职##大学生# 良田的田的微博视频
耳语III	#睡觉太显夜店视频#大学生##组合作业# 气温降低的温度是夏天的奇迹
adam亚当GUA	大学生给学校做了三亩四季花园#亚当#亚当设计的花园##亚当的农大花园#
三力sanili	四川文理学院yue了#大学生##五一假期去哪个演唱会#
黑眼圈小公猪00	清华大学请出战#高校#清华抵制#大学生##清华大学# 黑眼圈小公猪00的微博视频
高能冲浪指南	#这是什么新设计# 又活了一天#仲春欢乐启航# #大学生# 高能冲浪指南的微博视频
考古娱乐圈	广场如战场。原来大学生是这么整顿的 #吃屎##大学生# 考古娱乐圈的微博视频

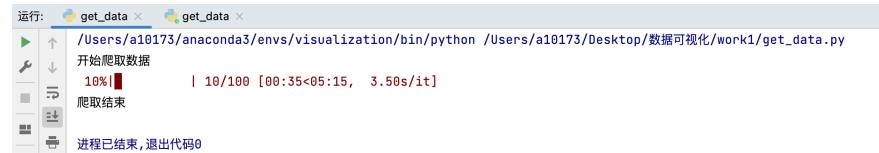
(2) 主题为“研究生”时：



爬取资料部分截图如下：(此为 2023 年 5 月 25 日爬取数据)：

研究生	
id	comment
首都经贸大学刘颖	今年硕士生就业情况比较好，我在首经贸读5名硕读硕士，2名去了国企；1名去了证券公司；1名去了事业单位；1名去了四大。我在社科院读1名硕士，考了公务员，#研究生##税务硕士##硕士研究生#
西大大妈	因为8岁了东华大学，大妈跟着逛就发现东南各个班，有娃的年妇都懂的，就算娃上不了，妈妈对这类事情也可有兴趣了。于是大妈搜了下，搜到了全国一共有三个少年班（不排除有更多的大妈没搜到），分别是中科院#研究生#一天的答卷，跟评估专业研究生又一届毕业了，祝贺你们
乌山学长	五一假期之后可能会和钱老师@钱婧BNU 一起做一期视频（主要是我提问，老师回答的形式），各位正在考研的朋友和即将读研的朋友有任何想读研考研方面的问题都可以评论区提问一下子，回头大家一起交流！#考研#24级
蕙蕙当庭	答客问，授予结果，研究生阶段我觉得就要结束了。晚上一个人坐着散散步，哼起了歌，我唱歌最难听，但我今晚很想唱《那些花儿》，这三年我从学校得到太多，她包容的、自由的、开放的、人文的一切滋养着幼稚的我
钱七七77	•••日本留学生活的不顺利去日本读读吧！！！太感谢你的！先介绍一下我的情况，我本科就读于日本立命馆大学，GPA排名大概是800左右，GPA为4.14/5，但是我也申请到了QS排名前15的英国爱丁堡大学的研究生项目，关于师弟的课题组喝啤酒的水桶倒掉#ning凌波事件以后遇到过这事请大家第一时间及时报警。有些小天使们嫌麻烦才不报警，#这么久# #研究生#
每天都发疯fff	考研的元气满满
中行杨桢	马克思主义哲学与理论文粹作者是荣春，南京航空航天大学人文学院法律系教授、法学博士；马光远，海南大学法学院博士研究生。来源是青苗法鸣#法硕##法硕##研究生#论文#
西湖管小理	西湖2023届硕博研究生学业奖：祝福大家前程似锦，开心快乐！XJTLU 西交利物浦大学#西交利物浦大学#校园#新星v计划##毕业季##硕士##博士##苏州##校园vlog##研究生# 苏州 西交利物浦大学 西浦管小理的微博视频
施老师求大成思政	公示期终于过了，我也是#研究生#了嘉兴 嘉兴学院(林校区) #考研#保科学院#马理论考研#嘉兴学院#
养真鱼的张宜良	别人985毕业靠实力，我在985毕业靠养真鱼#养殖#研究生##海王张宜良养鱼计划##出圈吧大学生##闪闪发光的同学##我在985大学养鱼# @微博校园 养真鱼的张宜良的微博视频 青岛·中国海洋大学(鱼山校区)
小董董老师	冒充导师给学生发这种短信，是什么奇葩逻辑？#研究生##考研复试#
土壤大叔	留学申请季，之前给小伙伴说了一波澳洲一年制的一些研究生专业！今天咱们邀请到了一位在澳洲国立大学Master of financial management硕士专业姐，分享出国留学就读体验#研究生##澳洲国立大学##出国留学##留学申请季#
别叫我奶盖仔	双录取985研究生的一天一起来投票#毕业季##研究生##中科院技术大学# @中科院技术大学 武汉·中科院技术大学 叫我奶盖仔的微博视频
别叫我奶盖仔	双录取985研究生的一天一起来投票#毕业季##研究生##中科院技术大学# @中科院技术大学 武汉·中科院技术大学 别叫我奶盖仔的微博视频
VinceLZH	做100次实验ELISA，满脸开心，藏都藏不住！#VinceLZH#同文的日常##研究生# 北京·清华大学郑裕彤医学楼
99本人	组织今天#考研#研究生#科研#上岸# 深圳·深圳大学丽湖校区
晶努力	23考研，一战上岸，双考上岸同次初试第三，总分第二上岸院校记录我的2021-2023，那段披星戴月、脚踏实地的日子。个中辛苦，都在字里行间了.....#考研上岸一路生花##考研上岸##双非考研##双次考研##考研#一战成绩##
中华美女博杨光	大数据把我推给每一个年龄焦虑的女孩！年纪只是数字，自信是照亮生命的灯塔，不被年龄束缚，坚信自己可以活出自己想要的样子！ #海外##研究生##离婚##全职妈妈##职场##外貌焦虑##年龄焦虑# 中华美女博杨光的微博
云南民族大学	#小伙伴们#研究生#云南民族大学祝大家复试顺利！！！都冲鸭！！！云南民族大学2023年硕士研究生复试考生须知网页链接云南民族大学2023年硕士研究生复试录取工作实施方案网页链接 昆明·云南民族大学
安心内哪	当然拒绝了导师老师的饼啊！上面把大学生#研究生#安安心哪#安心内哪的微博视频
Sunset-LJJ	大家好 这是我亲哥哥隆星宇是中国地质大学武汉研三的学生 今年24岁 因为突发性肺栓塞引起心脏骤停导致大脑缺氧，现在还在ICU一直昏迷不醒 我希望你们能够帮我多宣传和转发希望更多的人看到 救救我哥哥求求大家了 不求羊 走过的路每一步都算数了这次的带薪给自己！#考研#研究生##考研上岸#
扬州大学考研宿舍	扬州大学考研宿舍~扬州大学2023年硕士研究生录取名单正式公布 再次恭喜一声上的宝宝们👏👏 #研究生##考研#
瞄瞄orange	真的是气不过所以要吐槽，希望大家能看到这条微博的人都进退江南大学，研究生了，天天手机闹得早起叠被子，不然扣分影响评优评奖学金，更离谱的是我昨天收到辅导员消息说床垫厚了让我整改，我一个无语住，看研究生退宿#研究生#宿舍一角#关注网友天伦之乐#我好恨自己今天是何等？苦曰天临5年天伦5年春，我熬了四五个大夜，寝天微博也来得快一下好运#研究生#上岸##考研上岸##祝福研成功上岸# 福州·福建师范大学(旗山校区)
丽丽三品	感谢一路上的所有人 感谢我的主下一站在更加灿烂的道路上#研究生#四川师范大学 萍台
萨菲娅1	感谢一路上的所有人 感谢我的主下一站在更加灿烂的道路上#研究生#四川师范大学 萍台
江南大学考研宿舍	江南大学考研#研究生##决战考研# 请坚持下去，即使天荒地老也要永不放弃！
云龙水彩	电影有删减版，水彩有局部版#见艺#见艺#绘画艺术#绘画#画面#水彩#每一面#手帐#胶片#热爱生活#武汉##研究生# 武汉·湖北经济学院
墨染大腹马晋隆	如果你也像我一样拍的照片很崩那我应该觉得这个事情牵扯到马晋隆#研究生##我的春日Good#
光年速度飞	总有些惊悚的际遇，比方说当我遇见你#青春##美女##轻松一刻##亿点爆光计划##研究生# 光年速度飞的微博视频
Elavender	硕士进入1/2 wb 先来清一派正经毕业照库存#香港##毕业照##研究生# 香港·香港中文大学
好讨厌YanYan	#研究生# 午后好 上海·江川路地铁站
什么都管333	大学食堂卫生你不关注，教师师傅你不关注，校园霸凌你不关注，你倒是先切断普通人们对知识的渴望，你究竟在为谁谋福利，你究竟代表谁的利益，真的失望。#研究生#霸凌##考研网课##考研网课打假#
4Afamily	#我在大学待你##研究生#双战考研#如愿以偿！！4夺冠，我上岸4am战队 怪化
中国教育在线	#马云称都考研究生是严重浪费#【马光远】：人都考研究生，对解决就业没用处！】近些年，伴随着就业压力增大，考研热持续升温，分数也连年上涨，乃至本科学考研究生也很丢人一样。针对这种现象，著名经济学家马光远
一斐	开一个研究生专题，分享给我的研究生朋友们以及即将读研的朋友们：#读研之前必读的院士演讲：网页链接# 如何在读研的道路上快速失败：网页链接# 如何找到100篇关联文献：网页链接# 从导师角度看，赶dd的建议：#荣誉洞见时间# 大家平时阅读文献的时候有什么好的方法或建议呢？#studycount# #读书心得#
柒柒睡了	钱她的QA老师，关注您很久了，最近还在大忙还来请教你一下吧：双非研究生大研二读土木，考过sci综述（—作MDPI），中核有三篇，专利有六个，研二开始到计算机学院某大牛组做RA想直接硕转博，但现在看来老板想让我说起出国留学，醉划算的应该是年制硕士+研究生了！但是呢，很多同学都只能读英国的一年制硕士，却不知道澳洲的 哪们就给大家盘点澳洲的一年制硕士都有哪些：法学，教育学#研究生##英国硕士##新南威尔士大学##留#typerun
PTE易工炸	#研究生#论文#不仅读学生还要读导师# 我其实还在这些专业带研究生，希望能有真正有科研志向和热情的生选择我做导师。
用户7792342730	2023考研录取#录取# 请录取后这10件事一定要做到的都能被录取#23考研##考研打卡##研究生#
南京航空航天大学考研宿舍	南京航空航天大学考研#考研复试问题集锦！我不信你看完还不是研究生 见喜上岸！#南航#考研复试##研究生#
港深DE	近日香港留学十分火爆，竞争十分激烈，部分内地留学中介机构推出“保录”服务，从而向内地学生收取“中介费”中介费，更有声称200万元保录香港大学。港大今日发布严正声明，强调此事绝无可能！港大没有教授内推非本
留学申请季，英国澳洲硕士研究生要求雅思7.5，留学分别在语言分这一步，雅思转PTE，来围观学霸PTE听力口语90，从七做到八真是如何备考PTE的#澳洲硕士##offer##英国硕士留学##研究生##英国硕士# PTE易工炸的微博	山西经济倒数，除了政商环境极坏，行政效率低下，还有最大的问题的关于未来人才。就是高等教育的落后！严重影响后！山西招生考试网##研究生##山西考研成绩#一个考研成绩到现在都查不到了，全国独一份。公告也没有

(3) 主题为“杭州”时：

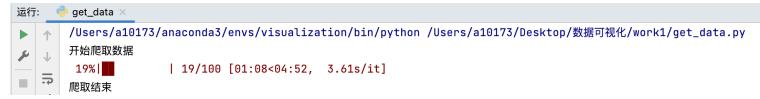


爬取资料部分截图如下：(此为 2023 年 5 月 25 日爬取数据)：

杭州

id	comment
叶艾茂	#杭州女子捐款1000万#，有钱又有爱，这回报社会的方式挺好！#外婆给慈善组织捐了1000万#杭州#
萤火虫如我	美国电商巨头亚马逊突然宣布，自7月17日起，停运亚马逊中文官方网站（Amazon.cn），与此同时，亚马逊中国将不再提供应用商店服务。报道称，亚马逊发言人表示，亚马逊中文官方网站停运前，公司会与卖家密切合作，确保
嘉木谦禾日当午	#杭州# 今日份打卡，周三开始啦，一周过得好快，新的一天一起加油，愿一切都顺顺利利！！奥利给，大家早上好
财经飞说不可	这些人都是有能力为楼盘垫底的#房价#杭州#
设计师-李功	跃迁设计 CHOW住宅：“即便是多”#室内设计# #家居设计# #杭州#
苟婵婵	跟杭州有关的记忆就是咧嘴笑😊——梦醒时分，杭州再见#杭州#舞剧红楼梦
陈雷Sammy	#浙江通报男子参加同性恋聚会#性学专家@李银河 曾提聚众淫乱罪，称这是一个过的错误的法律。来看看李老师的理由↓你支持吗？#杭州# #李银河建议取消聚众淫乱罪#
金燕lyncey	杭州的兄弟，你被绿了，希望你能削到泰国回来碰见一对露水鸳鸯，帮兄弟们避避雷，以下为信息，自行核对：你女朋友工作应该是策划一类。乘坐的是春秋航空9C8512清迈飞上海。机上你女朋友向旁边男的详细介绍了你的家
白水加VC	今天美丽的夜#杭州#夜色#
杭州巴士资讯	【#巴士风采#】杭州#都說杭州公車是慢浪漫的今年5.20不仅派出了两台“大红豆”双层巴士用于520特别活动，还有一辆与中国香港插画艺术家Ton Mak旗下FLABJACKS潮流品牌联合打造的“520膨胀巴士”，在杭州各大银泰商场车让人明明白一件事。“享受过程”和“自我感知”都很重要。OS. 有个爱骑行的男朋友真的很腿#饮饼旅行,三餐四季##杭州# 杭州
like一只菠萝	杭州锣鼓艺术舞蹈团除了给孩子良好的身体、优雅的气质还有根植于内心的价值观和不畏艰险的执着#杭州锣鼓艺术#随手拍##杭州#舞蹈##杭州舞蹈培训##你好五月# 跳雪寻梅李雷的微博视频
踏雪寻梅李雷	杭州的兄弟，你被绿了，希望你能削到泰国回来碰见一对露水鸳鸯，帮兄弟们避避雷，以下为信息，自行核对：你女朋友工作应该是策划一类。乘坐的是春秋航空9C8512清迈飞上海。机上你女朋友向旁边男的详细介绍了你的家
王金金WJJ-	#哈尔滨# 站圆满结束，非常感谢大家到场下一场：#杭州# 6月17日（刚好陪我过生日）拍照：小浩（自己的粉丝）做娇脸妆造：@焦京京_（我自己姐妹）做娇脸给哈尔滨嘉宾：高音宝宝真的太瘦太可爱了想抱抱呜~下一场比赛意大利老爸第一次看中国夜市，没想到这么晚，还有这么多人！#路卡和瑞丽#老外在中国#中国媳妇#武林夜市#老富#杭州#意大利人老富#夜市# Rachelle瑞丽的微博视频
Rachelle瑞丽	-Lyt-
MALLSTARS官方微博	#杭州地铁# 根据已批复的杭州地铁第四期建设规划，重新设计了杭州轨道交通2028年线路图，并拿出一张杭州都市圈绍兴、海宁、德清的线路图，其中绍兴二期规划的部分尚未批复。新图设计相比以前更加紧凑，希望大家喜欢
叉烧不是肥肉	#mallstars国际齐舞大賽#mallstars国际齐舞大赛#杭州#一派芬芳荷香溢绿韵撩纱舞霓裳爱在浙里，梦回杭州#7月23日，杭州分站赛我们又回来啦！以下为裁判阵容：ELEVEN老师@IMB-Eleven，Bad Machine老师@Badd_M
城探轩轩	#杭州# 跟球送老爸爸准备了一个热闹奥地利一带的市区感越来越棒在这样的场馆看球赛氛围不是我喜欢的都市联动感哈哈#杭州#杭州奥体中心体育馆
西安南瓜	#杭州# #深圳# 前两年去杭州 深圳接盘二手房的，现在看基本都亏麻了。——深圳杭州堪称全民炒房。深圳限制杠杠，立马打回原形；杭州供应太大，买二手如同进了堰塞湖。
Moon_羽儿胡儿	#宁波# “心心”的杭州#杭州#
杭州紫云仙院	吕祖这普通道论，对道家的修炼方法，给出了很精辟的论述，把修炼至命的内丹功从筑基，炼己到金液还丹自始至终的过程尽述无遗，是千百年来学道的一个指南，为学道之人奉为炼气、降丹之主旨。炼功必背哦！！！《吕祖论》#杭州# plog全程吃饭吃，竟然都没觉得虽然有些口讲确实不吃但也不算美食荒漠啊我感觉？打卡了两家比较火的店都还不错缘居是真的好吃（不过环境一般计划下次去吃金猪！plog#plog#日常碎片plog#plog博主日记##杭州#摄影# 把每一天的生活变成自己喜欢的频道#随手的#探索##杭州#
一朵花姑	05/02 05:00:00： 5月第二天的日出（昨天天好暗）#五一这样过#杭州#日出#杭州
喵喵-安然	今天看到离谱的遛狗者#杭州# #遛狗# 求生声带：不快真，如果是假的立刻刷。如果是谎言就太好了！欢迎指正，我立刻道歉。不接受上来就骂街人生攻击。纯吐槽，不表示任何立场
琳琅Ring	#好影像不怕黑# 今日五一假期后，杭州成为最热门的旅游目的地之一，接待游客超600万，尤其是西湖，游客到了摩肩接踵的程度。而出来旅游，最重要的就是发朋友圈，俗话说“没朋友圈的旅游是不完整的”，要想发好友圈
李子阳	来个“KISS”❤️#杭州# #随手拍# #街拍# #街拍日记# #摄影# #有哪些最美的人的景点#杭州#摄影##我的摄影日记#有杭州天目里
英嫂摄影inpetphoto	05/02 05:00:00： 早晨六点的西湖边#杭州# #遛狗# 求生声带：不快真，如果是假的立刻刷。如果是谎言就太好了！欢迎指正，我立刻道歉。不接受上来就骂街人生攻击。纯吐槽，不表示任何立场
张力视觉	#好影像不怕黑# 今日五一假期后，杭州成为最热门的旅游目的地之一，接待游客超600万，尤其是西湖，游客到了摩肩接踵的程度。而出来旅游，最重要的就是发朋友圈，俗话说“没朋友圈的旅游是不完整的”，要想发好友圈
Christin-小陳	来个“KISS”❤️#杭州# #随手拍# #街拍# #街拍日记# #摄影# #有哪些最美的人的景点#杭州#摄影##我的摄影日记#有杭州天目里
有个bong	#12岁少年植入人工心脏并开启充气人生# 【12岁少年变身“钢铁侠”植入人工心脏开启“充气”人生】24，在浙江大学医学院附属儿童医院的病房里，小患者成功完成心脏移植手术，术后恢复良好，顺利出院。他的外表看着与以后行程都发微博吧~哈哈哈哈哈~21号杭州站，偶遇吗？火线妹#火线妹##杭州#长沙
中航斯里	#西湖#西湖怎么来的#两千多年前的西湖，只是钱塘江边的一个小小海湾。耸立在南北两侧的吴山和宝石山，作为环抱海湾的两个岬角，在海水冲刷及泥沙积聚下，不断延伸，直至将浅海包围成内湖，这才有了最初的西湖。#西湖一大早，就去了灵隐寺，妈妈在网上查了时间，是七点开门，说赶早，于是6点多就出门了，从西湖边出发，一路美景，路过龙井村，湿润的空气里，仿佛有缕缕茶的茶香 多年前，拍摄《田教授家的二十八个保姆》，外景地在年轻人涌入杭州，杭州还要卷多久？#杭州#杭州网红#全景历史地理# 全景历史地理的微博视频
火线妹不是火箭筒	【#子弹壳羽绒服#040】#杭州# 要问湖里的鸳鸯鸟为何飞上枝头 要问戏水的鸭子是否知道水暖 要问为什么偶遇座莲没合身 要问夕照西下的杭州有多美 要问三月的，谁能不爱#杭州西湖# @杭州西湖风景名胜区 @摄影 人间仙境 #带着微博去旅行#杭州#西湖#
猴儿君的日常	#杭州# 早鸟奖 #在我看来公司愿意出这笔钱，员工愿意挣出这笔钱，公司最后的確也出了钱了，员工的確也拿到钱了，双方都诚实守信，挺好！管它是什么“早鸟晚鸟”还是“轮椅竞技赛”呢，一不违法二不违心双方你情我感就得了，师兄弟杭州一聚。合体武术与羽扇 #不负好春日# #运动就是坚持##传统文化#杭州# 武当杰克的微博视频
尚舍学长	#杭州#奶茶店使用霉变竹筒 【#杭州#杭州-奶茶店使用霉变竹筒】杭州奶茶店使用霉变竹筒，涉事奶茶店已停业。青岛生活娱乐的微博视频
陈小杏	【高丽寺·春雨海棠花】打卡美丽的中国杭州·西湖畔的古刹高丽寺，雨润海棠花！#遇见美好#杭州##摄影不凡#西湖#春游季#
地理公社	4月6日，浙江杭州的杭州湖滨步行街发生大变革，原来的方格走向被改为了“X”型。这一变化引起了许多网友的热议，有人说这样的景象就像电影里的场景一样，让人倍感新奇和兴奋。这种新颖的设计不仅增加了步行街国内外设计 印制 印制官网 X 土地土壤 网页链接 丘未茶研所 设计公司：@山地土壤室内设计 项目类型：#室内设计# #空间设计# 摄影版权：稳摄影项目地点：#杭州#
上海源渡的惬意生活	The Tree Room by@Worldland ▲在杭州去拜访了朋友新开的中古二手店，从线上到线下，从全球集萃好物到开始原创产品，很开心地把喜欢的事情做得有声有色，琳琅满目的宝贝清新轻松，天真浪漫，彰显着主人独特的审美
全景历史地理	#杭州# 我在等 星期五～ 杭州#湖漫枝
子瑜子	*半缕香水一城花。烟雨江南。*连续的阴雨天，让杭州的春天，从细腻柔美中，多了一份独特的清冽气质。虽然没有拍到阳光下的明媚春日，但一城烟雨的情景，也让我们感受到了古诗词中的氛围，飞鸟、落花、流水、雨滴、
乔木榆长	
文森诗画	
武当杰克	
青岛生活娱乐	
摄影视觉杨	
白水晶爱上	
印制美学	
SSEE-LAB	
容嬷嬷的姐姐容婆婆	
回形岛	

(4) 主题为“周杰伦”时：



爬取资料部分截图如下：(此为 2023 年 5 月 25 日爬取数据):

周杰伦	
id	comment
嘻笑堂HipHop	过去一幕幕浮现！周杰伦《轨迹》舞台！“我会发呆，然后忘记你”#周杰伦 #周杰伦上海# 你是哪年开始听周杰伦？ 嘻笑堂HipHop的微博视频
周杰伦国际后援会Jay2u	周杰伦 微博 #周杰伦#
琪儿Leo	香港之行Vlog 来啦！吃吃吃 ✅买买买 ✅杰伦演唱会 ✅都完成！超级无敌开心的假期～琪儿Leo#电子秤菜#香港##周杰伦# 琪儿Leo的微博视频
陈一鸣lan	那些年我们一起追过的周杰伦P.S.夜曲，你听得到，我不配。她的睫毛，一路向北还有蒲公英的约定就留给下次吧～#周杰伦香港演唱会##周杰伦# 陈一鸣lan的微博视频
鬼才bb出逃记	我会笑着 然后忘掉你～#轨迹#轨迹#音乐#周杰伦# 鬼才bb出逃记的微博视频
环球音乐版权	祝贺环球音乐版权代理词曲作者及作品在第34届金曲奖入围17项大奖！完整名单一键查看～♪ 🎵 @吴青峰 及作品斩获最佳作曲人、最佳华语男歌手奖、年度专辑奖、最佳华语专辑奖等7项提名。 🎉 @Matzka马斯卡 及作品，高龄#周杰伦 今天是个特别的日子，竟然偶遇到我童年的男神，周杰伦，我真的很开心。#亿点曝光计划# 高龄班的微博视频
高碧瑶	地上捡到票。我喜欢的世界末日轨迹都唱了。只不过没敢去前面要签名我是不是第一个带five guys 进去周杰伦演唱会吃的人#这个很费劲##周杰伦# 网页链接
RealF费曼	终于轮到我来把这个演唱会开场了！#周杰伦#演唱会#周杰伦香港演唱会#周杰伦 大看台
大看台	网友：这只是场馆外的极限，不是周杰伦的极限。#周杰伦香港演唱会##周杰伦#
周杰伦歌迷网官方微博	网友：这只是场馆外的极限，不是周杰伦的极限。#周杰伦香港演唱会##周杰伦#
每阅读	🔥周杰伦专辑封面#周杰伦#超话#周杰伦#周杰伦演唱会#周杰伦#周杰伦演唱会现场点歌#雨下一整晚#完整版#周杰伦#王俊凯#周杰伦音乐速递的微博视频
周杰伦音乐速递	周杰伦香港演唱会第六场王俊凯现身周杰伦演唱会现场点歌 #雨下一整晚# 完整版#周杰伦#王俊凯#周杰伦音乐速递的微博视频
周杰伦的小喇叭	周杰伦的人气究竟有多恐怖，这只是场馆外能承受的极限人迹，而不是周杰伦的极限。#周杰伦香港演唱会##周杰伦#周杰伦演唱会#
JoeyThyo戴祖恒	幸福的氛围~#杰伦子爵#周杰伦# @JimmyNg_吴若希
QQ飞车	周杰伦 #周杰伦# zj周杰伦拍历代电影时的花絮混剪！杰伦真的又帅又可爱啊！好喜欢周杰伦啊啊啊！B站视频号指路：BV1eh411c7Si QQ飞车的微博视频
周星驰影迷会	#星ins# 他们除了卖酒，什么都可以合作#周星驰 周杰伦#
孙铭维	今晚杰伦演唱会喜欢瓶子和胡椒，点了几千里之外#周杰伦#周杰伦香港演唱会#周杰伦香港#香港中环海滨活动空间(香港中环劳道店) 孙铭维的微博视频
阿呆民	#董大尉# #翰赵# #周杰伦# 他们用什么语言解释这种感觉.....
王小草Nova	没有想象中的崩溃大哭，没有歇斯底里，没有开心到癫狂。那一刻我只是在想，我一定要记住这张脸，一定要记住现在这个时刻——他在舞台上的样子。甚至期待他90岁还能站在舞台上，肯定和今天一样，帅爆了！有很多想说#周杰伦#记得不能说的秘密呢？还记得叶澜伦，路小小雨吗 #周杰伦#
爱唱歌的小T	惊喜！周董和周杰伦#强势单曲《电影》超强mv大首播！Houson猴姆的微博视频
Houson猴姆	惊喜！周董和周杰伦#强势单曲《电影》超强mv大首播！Houson猴姆的微博视频
夏城南	#无声卡清唱##周杰伦# 我不落泪，忍住感觉。夏城南的微博视频
娱营	#周杰伦#感谢！这是02年啊！他居然就已经把演唱会开成这样了 the one的唱号，live的天花板 烂营的微博视频
周杰伦资讯源	#周杰伦# 低调 MV 紧+简体 1080P
孙伊_Selina	周杰伦演唱会vlog#谁能把拒绝下班之后去看周杰伦演唱会呢#周杰伦香港演唱会##周杰伦#周杰伦演唱会# 孙伊_Selina的微博视频
美酒de微博	今晚，是属于我们的杰伦的！！#周杰伦# #周杰伦#演唱会不会
问问呀	真的做梦，是从小听到大的偶像周杰伦香港演唱会##周杰伦# 香港-中环海滨活动空间
JXR-雾气泡水	王俊凯去香港看周杰伦演唱会，与偶像合唱《雨下一整晚》，一开口就是老歌迷了，小飚是周杰伦的超级铁粉，等这一天已经等了24年～#周杰伦#王俊凯# JXR-雾气泡水的微博视频
没有了以后Zz	周杰伦#周杰伦香港演唱会# Day7#周杰伦#现场尽享 solo好歌秀 #周杰伦# 演唱会#没有了以后Zz的微博视频
周杰伦屡遭逮	#王俊凯#香港精周杰伦#王俊凯# 今年现身看周杰伦香港演唱会，点歌《雨下一整晚》，并合唱《雨下一整晚》杰伦：哎哟，欢迎王俊凯。杰伦说：是不是从小听我的歌长大的，王俊凯回答：对，我等这一刻等了24年。（视频#溥仪眼镜全力支持，「保诚保险用心聆听」#周杰伦#周年世界巡回演唱会2023 - 香港站，经已圆满落幕！台湾音乐精灵王周杰伦，多场精彩演出，带来多首经典金曲，全城掀起热潮！让我们一起重唱演唱会精彩时刻。#
溥仪眼镜	溥仪眼镜全力支持，「保诚保险用心聆听」#周杰伦#周年世界巡回演唱会2023 - 香港站，经已圆满落幕！台湾音乐精灵王周杰伦，多场精彩演出，带来多首经典金曲，全城掀起热潮！让我们一起重唱演唱会精彩时刻。#
恋爱动力学	王力宏#刘畊宏健身法# #周杰伦# 王力宏放周杰伦鸽子，还说这“急也跟着跳”很会挑时间 恋爱动力学的微博视频
潇舒克	我也是有To点的人！（不如今晚睡前说一句你最喜欢的AYAOHU歌词吧#周杰伦#
李俊毅-JY	昨晚的周杰伦全场大合唱终于有找一份子有在现场的朋友吗？@被逼营业的芳芳 差点下暴雨，还好天公作美#周杰伦# #周杰伦香港演唱会# 香港音乐狂想曲
音乐狂想曲	周杰伦跟着刘畊宏跳舞，在厦门拍摄周杰伦# #周杰伦#鼓浪屿# 音乐狂想曲#周杰伦#周杰伦的微博视频
盖世英雄玉橄榄	周杰伦能不能换个发型，感觉之前油腻头的好看很多，就算是刘涛，这个刘海我觉得也不适合他，年轻的时候的刘涛才比较适合。（非广告）#周杰伦#鼓浪屿#
悟空音乐随笔	#周杰伦刘畊宏一起跳本草纲目# 周杰伦刘畊宏跳《本草纲目》，终于同框了，杰伦私下没少练吧，动作熟练有模有样！#周杰伦#鼓浪屿# 悟空音乐随笔的微博视频
Music_Data	周杰伦没有多少鼓浪屿，但是接下来的演唱会还是会有很多的。#周杰伦#鼓浪屿#
SwapFuck	黄巧实力上线周杰伦又双収秀球技了！此时刘畊宏心情_____？#周杰伦#双収秀球技#周杰伦# SwapFuck的微博视频
目睹了一切的R9	周杰伦 跳操了解 😊 #周杰伦#鼓浪屿# #周杰伦#厦门#
海港SirIbaSa	周杰伦香港演唱会情况，周杰伦粉丝还是最长情的，二十多年坚持的喜欢一个人真的不容易！看香港演唱会好多粉丝站在外面连有坐在草地上和游艇上一起唱！想象着杰伦六七十岁在台上就算不唱歌我们一起大合唱将帅是怎样的第一次看周杰伦演唱会祖国反黑#香港警察#我是中国人#周杰伦#
大小眼睛	周杰伦#周杰伦# 不是调侃离山也不是放鸽子哦，是因为岛上歌迷太多了，为了安全起见取消了，然后改去了六中和英才学校，羡慕见到杰伦的同学们
JFOR_ME	#周杰伦#鼓浪屿# 不是调侃离山#周杰伦这次到了鼓浪屿，有杰伦的地方就是烟火，感受到粉丝的热情，去过两次，很美的地方，不愧是音乐届的顶流。吟墨香的微博视频
吟墨香-	这是周杰伦距离最近的一次#刘畊宏#周杰伦#周杰伦香港演唱会# 周杰伦#吟墨香的微博视频
呗呗兔	周杰伦香港音乐会最近的一次#刘畊宏#周杰伦#周杰伦香港演唱会# 周杰伦#呗呗兔的微博视频
周杰伦音乐资讯	高清图！周杰伦在五缘湾游泳会与刘畊宏一起跳操 😊 #周杰伦#鼓浪屿#
VIVISHU薇薇舒	哥太帅了 多赚点 💰下次去韩国看 #周杰伦#周杰伦演唱会# 周杰伦#
你是红...	关于周杰伦回我的高中开演唱会这件事#厦门六中#周杰伦#

4. 代码简介

代码目标是爬取微博上的帖子并将结果存储在 CSV 文件中。主要用到了 requests、pandas、re 和 tqdm 库。

首先，定义了一个要搜索的话题和爬取的轮次数。

然后，创建一个 CSV 文件用来存储结果，文件名根据指定的话题决定。CSV 文件的标题行为“id”和“comment”。设置 HTTP 请求的 header 和 cookies，这些信息将在后面的 requests 请求中使用。使用 trange 函数（tqdm 库的一个函数，用于在 Python 长循环中添加一个进度提示器）来遍历指定的轮次数。对于每一轮，构造一个 URL，该 URL 将话题和当前的页码插入到预定义的 URL 模板中。使用 requests 库的 get 函数向构造的 URL 发送 HTTP 请求，并将返回的 JSON 响应存储在 json 变量中。再解析 JSON 响应，对于响应中的每一条卡片（每条微博帖子），提取用户名和帖子文本，然后将这些数据存储在 DataFrame 中，并将其追加到之前创建的 CSV 文件中。这里用到了正则表达式来清除文本中的 HTML 标签。

最后，在每一轮结束时，休眠 3 秒，防止由于高频连续爬取而被微博服务器封锁。