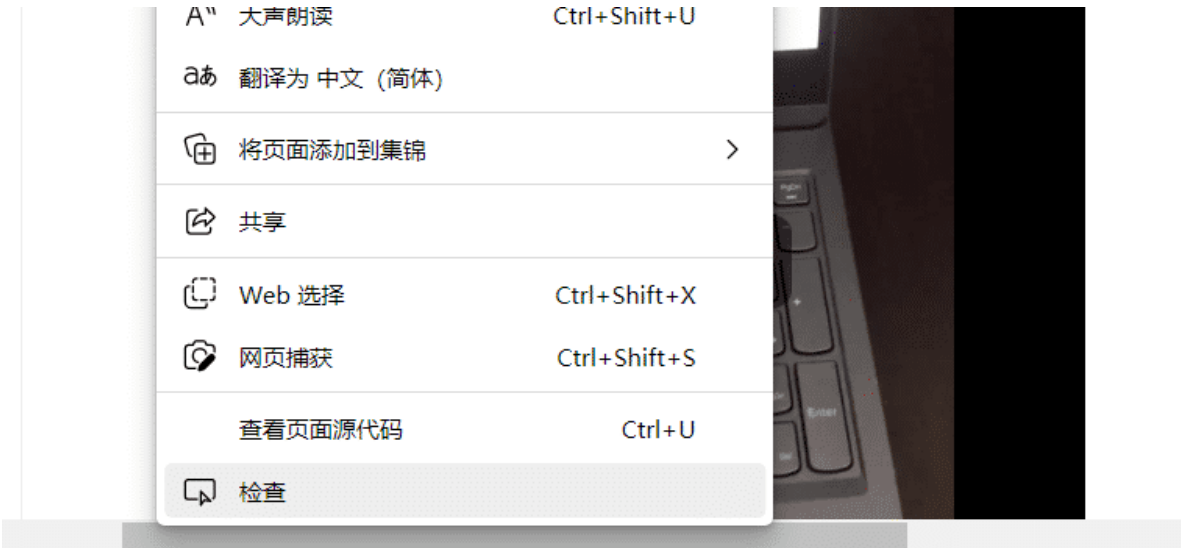


爬取京东评论

确定目标 找到正确url

右键打开检查



点击网络

 S***u

★★★★★

包装保护：2层纸箱，内有海绵包裹很安全。

外形外观：金属质感，低调的配色，个人很喜欢。

运行速度：运行流畅，3种模式任意切换。

画面品质：2k屏没的说，非常清晰，看起来很舒适。

元素 控制台 源代码 网络 x 十

保留日志 禁用缓存 无限制

Fetch/XHR JS CSS Img 媒体 字体 文档 WS

Cookie 已阻止请求 第三方请求

s	20 ms	30 ms	40 ms	50 ms	60 ms	70 ms	80 ms	90 ms

评论翻页可以快速找到正确的文件



找到请求url 以及 请求头

请求 URL: https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment98&productId=100019125569&score=0&sortType=5&page=1&pageSize=10&isShadowSku=0&rid=0&fold=1

请求方法: GET

状态代码: ✔ 200

User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/106.0.0.0 Safari/537.36 Edg/106.0.1370.37

鉴于京东网页的机制，需要通过请求头伪装成用户进行访问

初次尝试还是有些困难的，尤其是不知道自己找没有找对

代码实现

实际上还是“借鉴”别人的代码较多

导入包

request、json、time

都是常用的包

伪造请求头

```
headers = {'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/106.0.0.0 Safari/537.36 Edg/106.0.1370.37'}
```

可以自己从浏览器获取

请求部分

```
url = 'https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment98&productId=100019125569&score=0&sortType=5&page=1&pageSize=10&isShadowSku=0&rid=0&fold=1'.format(page)
response = requests.get(url, headers=headers)
```

通过改变page来改变评论页数达到翻页效果

数据处理+数据写入

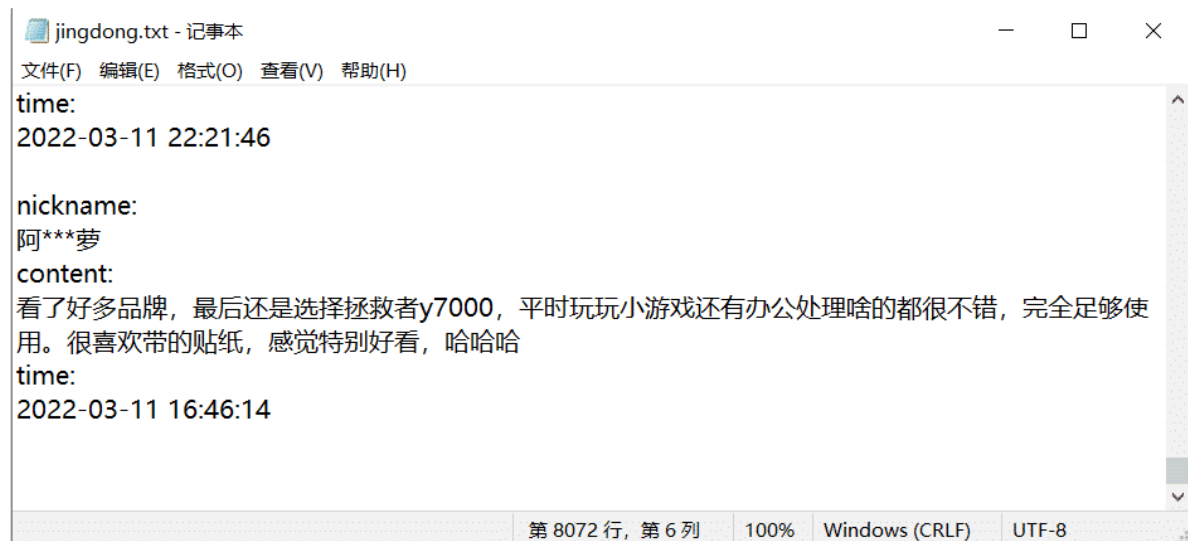
```
# 将不需要的部分 删除/替换 使剩下的部分成为一个列表
data = response.text
jd = json.loads(data.lstrip('fetchJSON_comment98vv12345(').rstrip(');'))

data_list = jd['comments']
for data in data_list:
    nickname = data['nickname']
    content = data['content']
    creationtime = data['creationTime']
    print(nickname)
    print(content)
    print(time)
    # 数据写入txt文件:

    fp.write(f'nickname:\n{nickname}\ncontent:\n{content}\nntime:\n{creationtime}\n\n')
    # 鉴于京东的反爬机制, 当一个ip短时间发起大量请求时会暂时性锁ip, 所以加入一个“sleep”限制爬取速度。猜测2s一页都是安全的
    time.sleep( 3 )
```

本地文件结果展示

本来是想写入csv文件的, 但是搞了半天没搞来



1k条评论

使用MySQL储存

软件配置

MySQL (本体) + sqlyog(可视化软件)

虽然说MySQL有自带的workbench可视化平台

但sqlyog的“选课”确实爽

录入数据

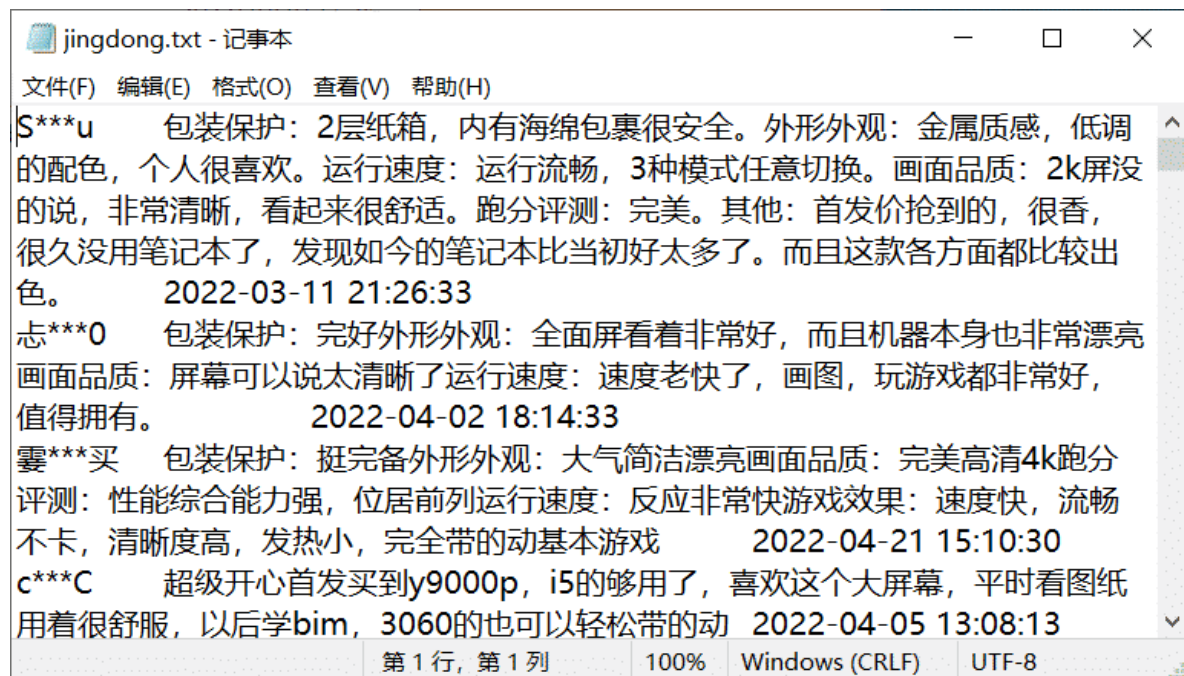
```
--准备工作
CREATE DATABASE comment;
USE comment;
CREATE TABLE `comment`.`comment` (
  `nickname` VARCHAR(20) NOT NULL COMMENT '用户昵称',
  `content` VARCHAR(500) NOT NULL COMMENT '评论内容',
  `time` VARCHAR(50) NOT NULL COMMENT '评论时间',
  PRIMARY KEY (`nickname`)
) ENGINE=INNODB CHARSET=utf8mb4;
--sql yog可视化创建表本质上还是对应所需的sql语句

--录入数据
LOAD DATA LOCAL INFILE 'C:/Users/王南钦/Desktop/22-王南钦-10.4-138/#6-王南钦/jingdong.txt' INTO TABLE `COMMENT`;
FIELDS TERMINATED BY '\t';
LINES TERMINATED BY '\n';
```

实际上，上图的展示格式并不能正常录入到MySQL的table中

更改格式重新写入数据到txt中

```
# 补充到for循环内
content = content.replace('\n', '')
# 重写
fp.write(f'{nickname}\t{content}\t{creationtime}\n')
```



虽然是丑了些，但只有这样才能被正确读入

再使用sql写入table+

<input type="checkbox"/>	nickname	content	time
<input type="checkbox"/>	青***c	包装保护：两层纸箱包裹包装很好~不会磕磕碰碰外形外观：外观很漂亮 边框很窄画面品质：看着挺好的 看视频画面很	2022-03-27 14:59:03
<input type="checkbox"/>	霖***子	包装保护：外观完好，无使用痕迹，确认不是摸摸机外形外观：摸上去质感很棒，不沾指纹，小键盘和滴血方向键很满意！画面	2021-12-12 09:27:45
<input type="checkbox"/>	韦***年	包装保护：双层保护外形外观：很漂亮，棱棱角角很分明，摸上去很舒服画面品质：画质非常好跑分评测：没测，但感	2022-04-17 14:21:38
<input type="checkbox"/>	顺***g	快递速度很快，包装保护很好	2022-05-02 15:41:52
<input type="checkbox"/>	顺***侯	包装保护：包装严严实实没有任何问题外形外观：很惊艳非常满意画面品质：没什么问题跑分评测：拿鲁大师跑了140多	2022-03-12 07:48:47
<input type="checkbox"/>	风***勾	物流非常快，包装的很好，电脑也很nice，心心念念了半年了，终于到手了，电脑很帅，很清晰很流畅（赞）！大家冲	2022-03-11 20:29:48
<input type="checkbox"/>	马***m	发货后一天就收到了，颜色质感都不错，开箱迫不及待的试了试，手感非常不错，游戏质感超级棒，漆色也很好，非常	2022-03-12 10:52:42
<input type="checkbox"/>	高***0	新款十二代处理器速度很快，外观都是金属壳，内存是DDR5-4800频率，打游戏都足够了，显卡3050Ti做设计游戏很流畅	2022-04-04 11:37:17
<input type="checkbox"/>	高***也	预售货足，不用抢购，很赞。电脑质感很棒，用来固定办公厚度正合适。屏幕和音效都不错。2k屏幕，雷电接口，DP充	2022-04-03 17:37:11
<input type="checkbox"/>	高***人	久闻联想拯救者大名，这次属于是付诸行动了，有幸抢到了首发的y9000p包装保护：包装很严实，没有任何碰撞，好评	2022-03-11 12:14:46
<input type="checkbox"/>	金***海	这是我在联想买的第二台机器，非常奈斯，相对于同价位产品，这个机器真的很棒，拯救者开机是真的快。以后还会继续	2022-03-30 12:13:28
<input type="checkbox"/>	鸿***8	机器性能非常强大，目前在玩的大型游戏都绰绰有余。而且还支持三档调节，日常轻度使用，网游风扇声音非常小，买	2022-03-27 23:30:09
<input type="checkbox"/>	聯***安	外观上金属材质，太空灰很有科技质感，我非常喜欢。三种切换模式，组合键切换方便流畅，总体运行特别快。屏幕选这款	2022-03-31 12:28:47
<input type="checkbox"/>	虞***舜	画面品质：画质清晰，细腻包装保护：包装扎实，完好无损外形外观：外形美观大方运行速度：运行速度很快跑分评测：	2022-01-28 10:24:07
<input type="checkbox"/>	凌***德	很好的机器。 玩网游没问题，各方面配置均衡 wifi6e网卡 三星内存 镁光的固态 没有短板	2022-04-02 00:26:49

大功告成！