

工具准备

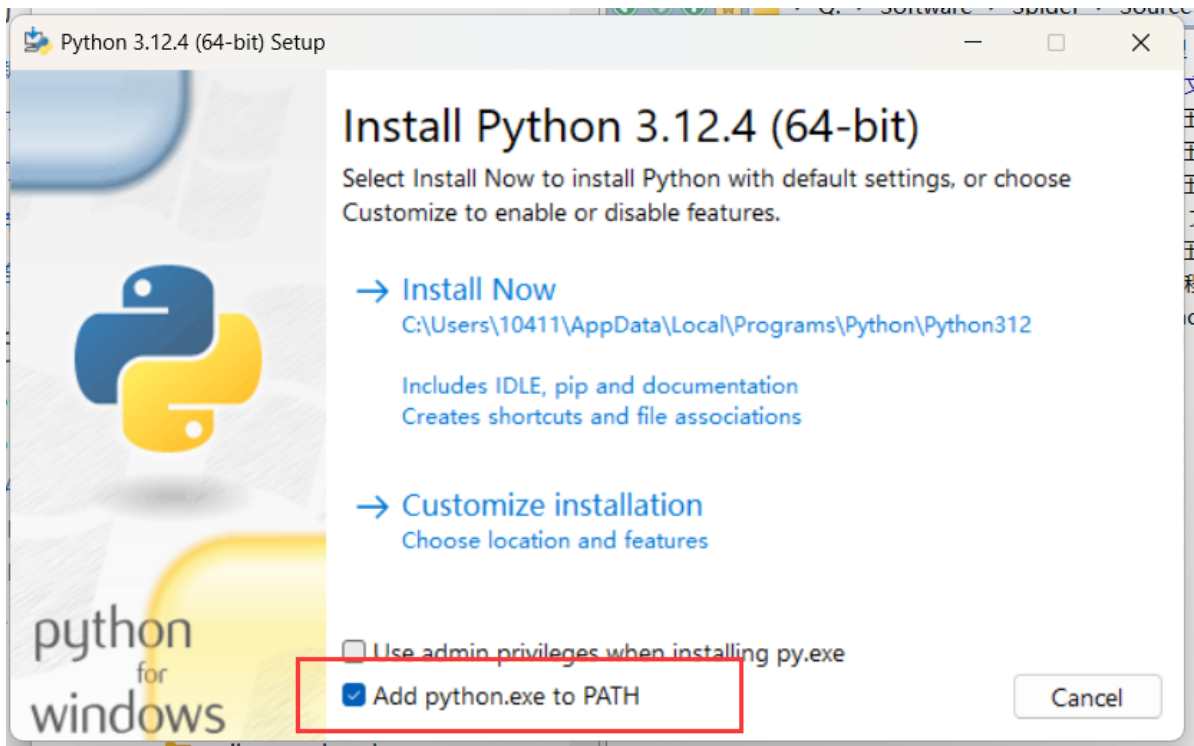
vscode



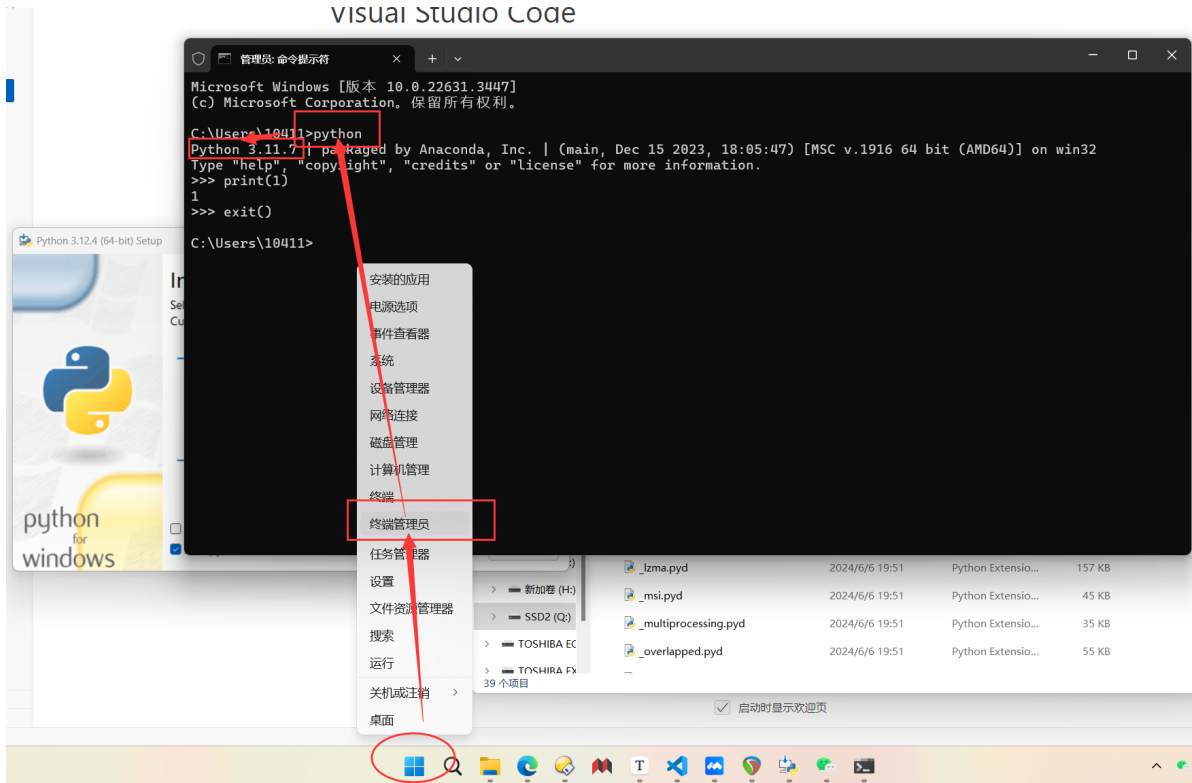
扩展安装Chinese、Jupyter、Python

code runner、Rainbow CSV、SQLite Viewer

python



安装并添加到环境变量，在终端输入相应命令，确认安装完成，如下图



mac这样打开终端，输的命令可能python改为python3，可以试一试

如何打开终端！

终端是系统中一个应用程序，你可以直接在所有程序中找到它，点击打开就行。我比较喜欢使用聚焦搜索，输入「终端」或「Terminal」，看到终端被选中了按下回车即可。



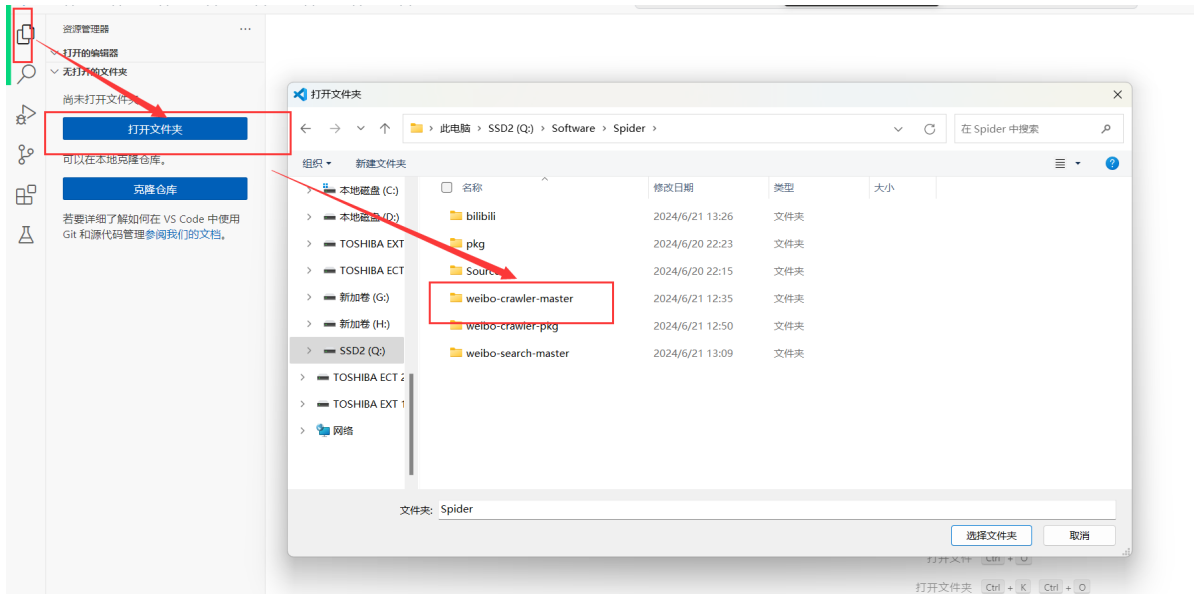
在vscode里也可以检查python版本

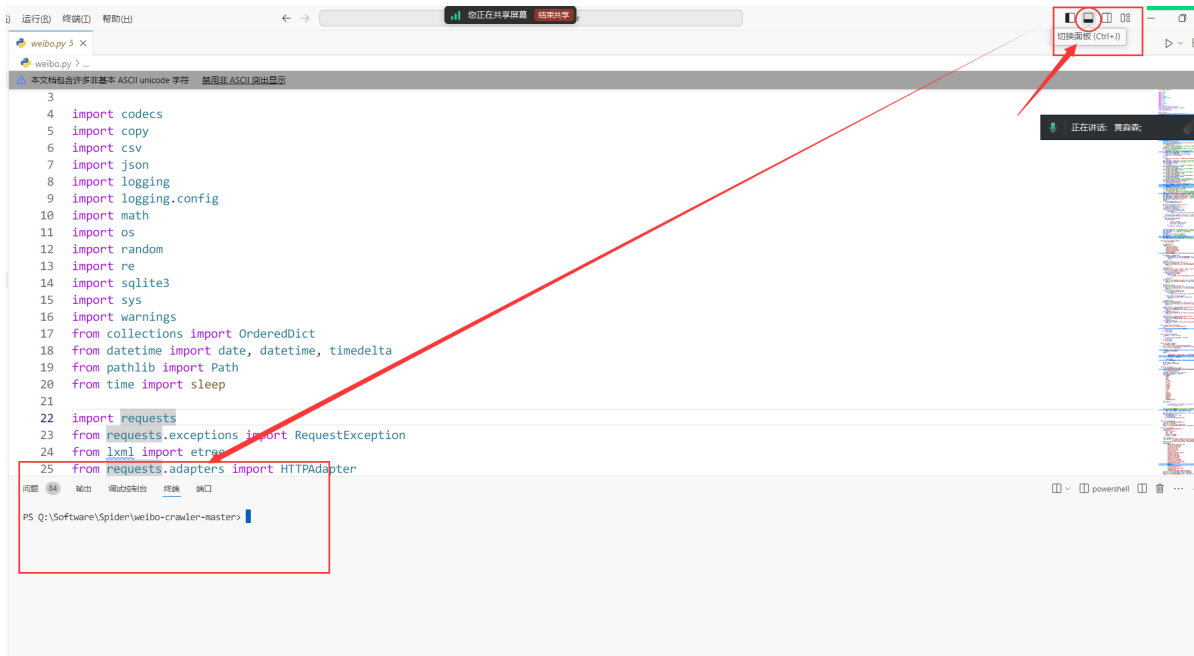
```
1 #!/usr/bin/env python
2 #-*- coding: utf-8 -*-
3
4 import codecs
5 import copy
6 import csv
7 import json
8 import logging
9 import logging.config
10 import math
11 import os
12 import random
13 import re
14 import sqlite3
15 import sys
16 import warnings
17 from collections import OrderedDict
18 from datetime import date, datetime, timedelta
19 from pathlib import Path
20 from time import sleep
21
22 import requests
23 from requests.exceptions import RequestException
24 from lxml import etree
25 from requests.adapters import HTTPAdapter
26 from tqdm import tqdm
27
28 import const
29 from util import csvutil
30 from util.dateutil import convert_to_days_ago
31 from util.notify import push_deer
32
33 warnings.filterwarnings("ignore")
34
35 # 如果日志文件夹不存在, 则创建
36 if not os.path.isdir("log/"):
37     os.makedirs("log/")
38 logging_path = os.path.split(os.path.realpath(__file__))[0] + os.sep + "logging.conf"
39 logging.config.fileConfig(logging_path)
40 logger = logging.getLogger("weibo")
```

微博爬虫工具

crawler

安装, 解压文件夹, 然后在VScode里面打开





在终端中一条一条输入，以确认安装没问题

```
python --version
```

在终端里输入，安装程序

```
python -m pip install -r requirements.txt
```

打开config.json设置软件（看README具体设置）

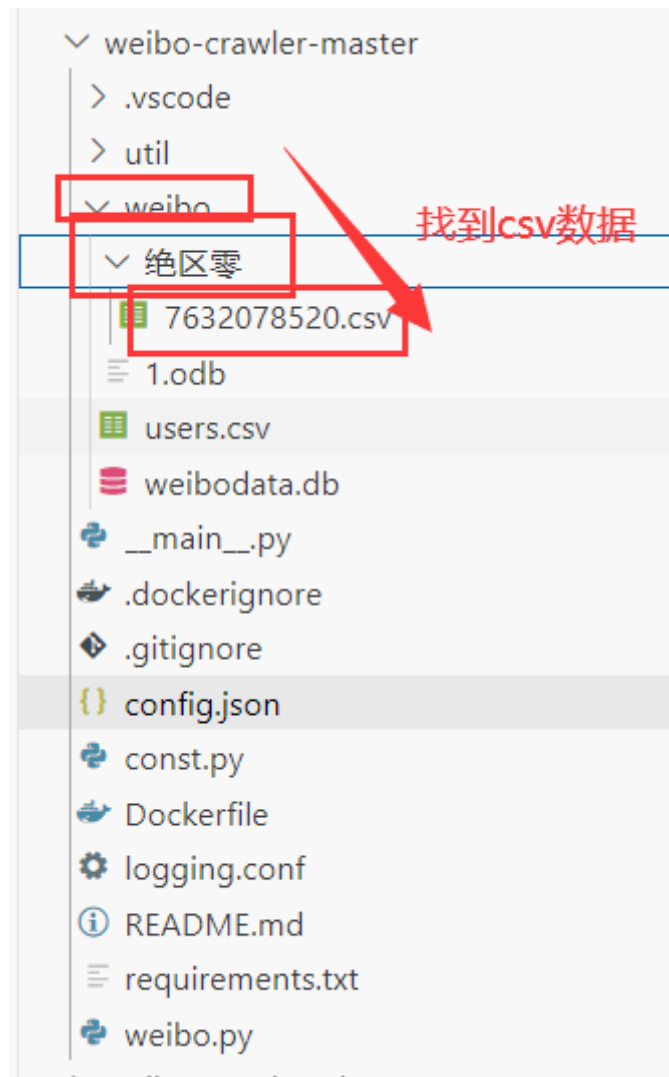
如果只需要爬取某个博主的博文

```
{  
  "user_id_list": ["7632078520"],  
  "only_crawl_original": 0,  
  "since_date": "2024-06-19",  
  "start_page": 1,  
  "write_mode": [  
    "csv",  
  ],  
  "original_pic_download": 0,  
  "retweet_pic_download": 0,  
  "original_video_download": 0,  
  "retweet_video_download": 0,  
}
```

博主ID

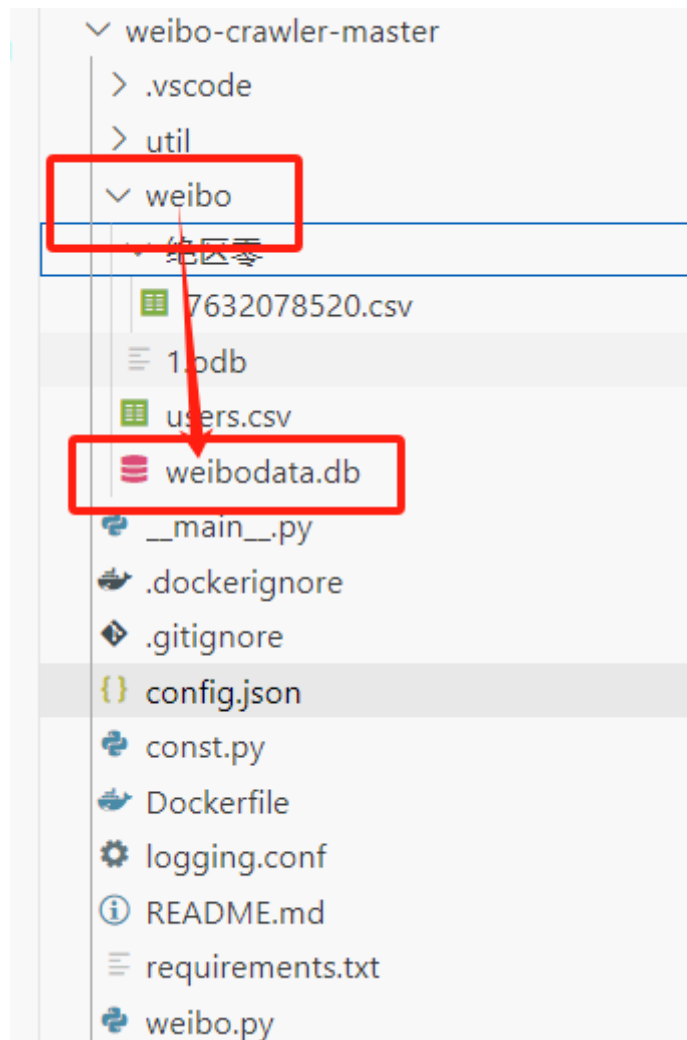
从哪天到现在

0意味着不
下载视频
图片



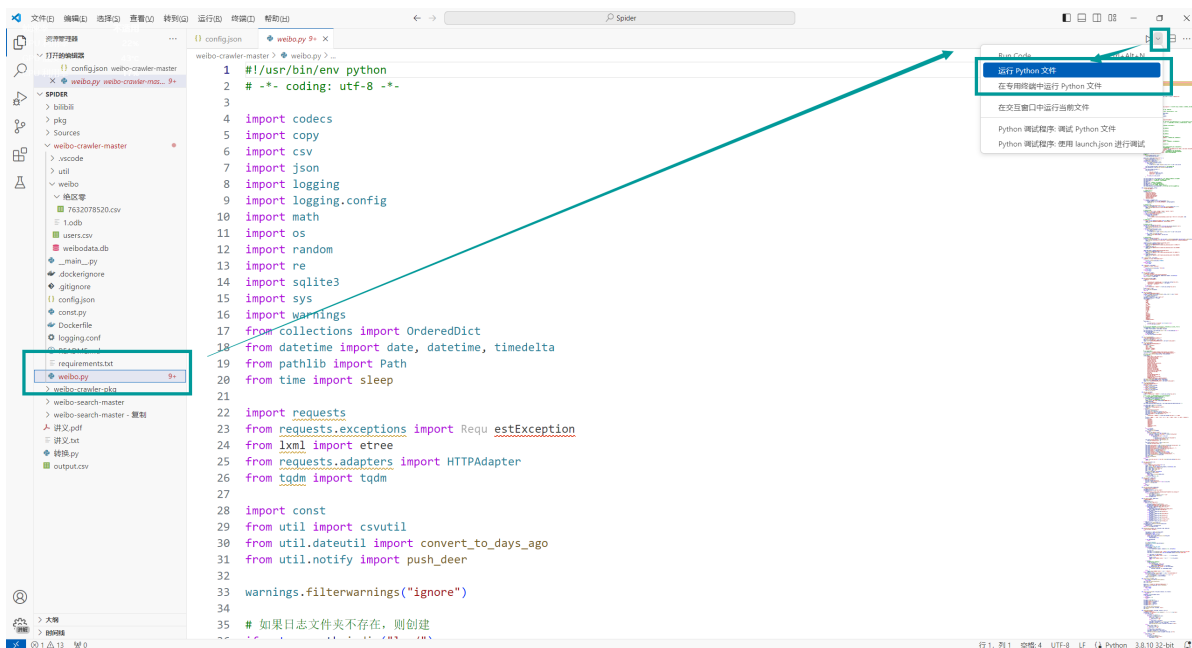
如果需要爬取博文下面的评论

```
1  {
2      "user_id_list": ["7632078520"],
3      "only_crawl_original": 0,
4      "since_date": "2024-06-19",
5      "start_page": 1,
6      "write_mode": [
7          "sqlite" 这个格式才能保存评论
8      ],
9      "original_pic_download": 0,
10     "retweet_pic_download": 0,
11     "original_video_download": 0,
12     "retweet_video_download": 0, 1表示下载评论
13     "download_comment": 1,
14     "comment_max_download_count": 20,
15     "download_repost": 1,
16     "repost_max_download_count": 20,
```



运行爬虫

终端里面运行weibo.py



数据库db转为csv

```
1 import sqlite3
2 import csv
3
4 # 把database.db改成文件路径(右键文件复制路径)
5 conn = sqlite3.connect('D:\Software\Spider\weibo-crawler-master\weibodata.db')
6 cursor = conn.cursor()
7 # table_name改为对应表名
8 query = 'SELECT * FROM table_name'
9 cursor.execute(query)
10 rows = cursor.fetchall()
11
12 # 在这里该生成的文件名
13 filename = 'output.csv'
14
15 with open(filename, 'w', newline='', encoding='UTF-8') as file:
16     r = csv.writer(file)
17     r.writerow([i[0] for i in cursor.description])
18     r.writerows(rows)
```

The screenshot shows a code editor with a Python script. A context menu is open over the file explorer, with '复制路径' (Copy path) highlighted. Red and green arrows point from the code to the file explorer and the database table view respectively.

The database table view shows the following columns: id, ext, data, weibo_id. The table contains one row with the following data: comments, response, sqlite_sequence, user, weibo.