Crawling bilibili

**1. Experimental environment**

**1.1 VERSION**

PYCHARM（COMMUNITY EDITION）2021.2.2

PYTHON 3.9

* 1. **Lib**

re

requests

time

pandas

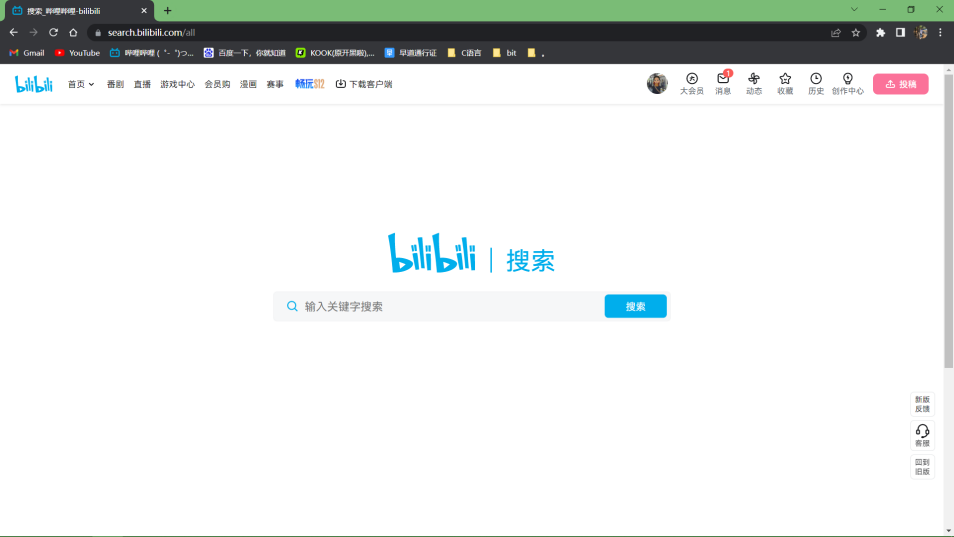
bs4

**2. Web page analysis**

**2.1 BiliBili search engine**

a) First, open the website for bilibili:http://search.bilibili.com/.

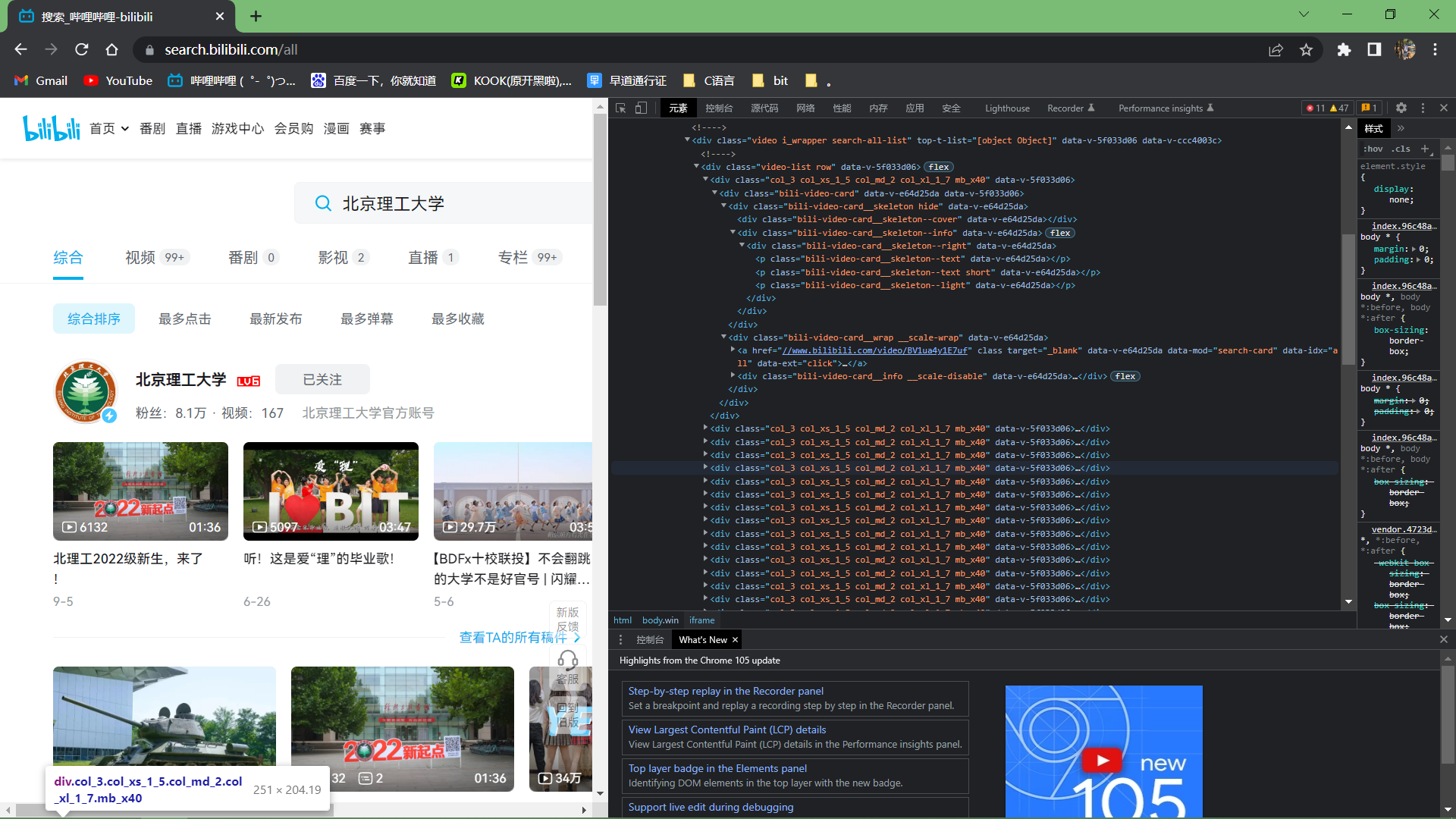
By input the keyword we can get search page we need.



b) Enter the kitchen knife to search.



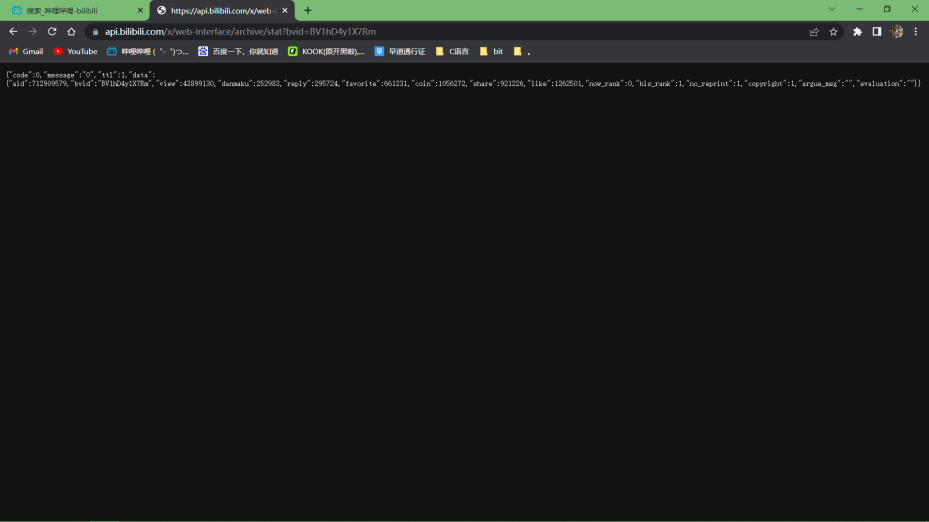
c) Find each video’s BV number



d) After getting the BV number we can find each video’s details

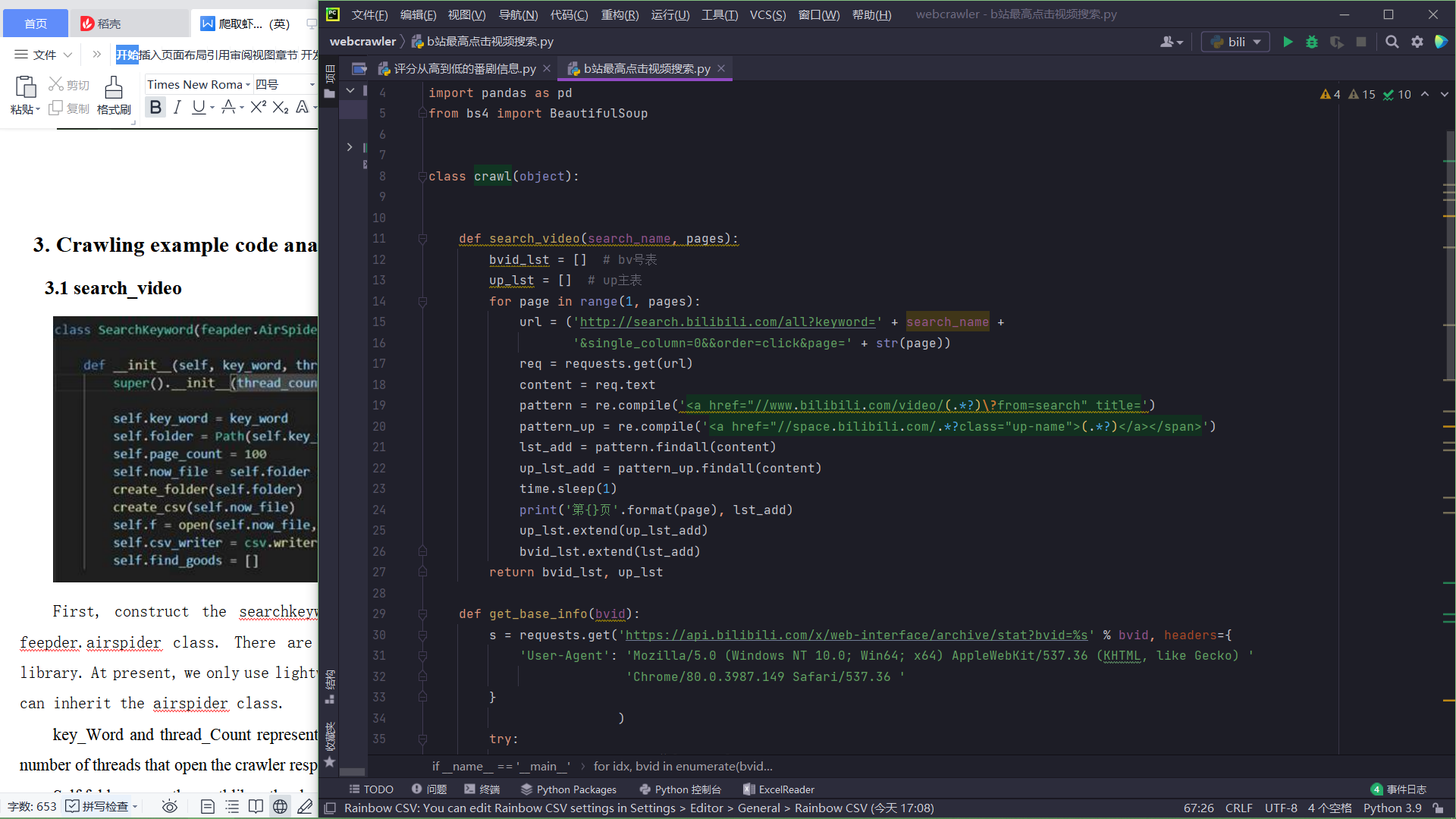
https://api.bilibili.com/x/web-interface/archive/stat?bvid=%s' % bvid

e) Analyze the JSON file.



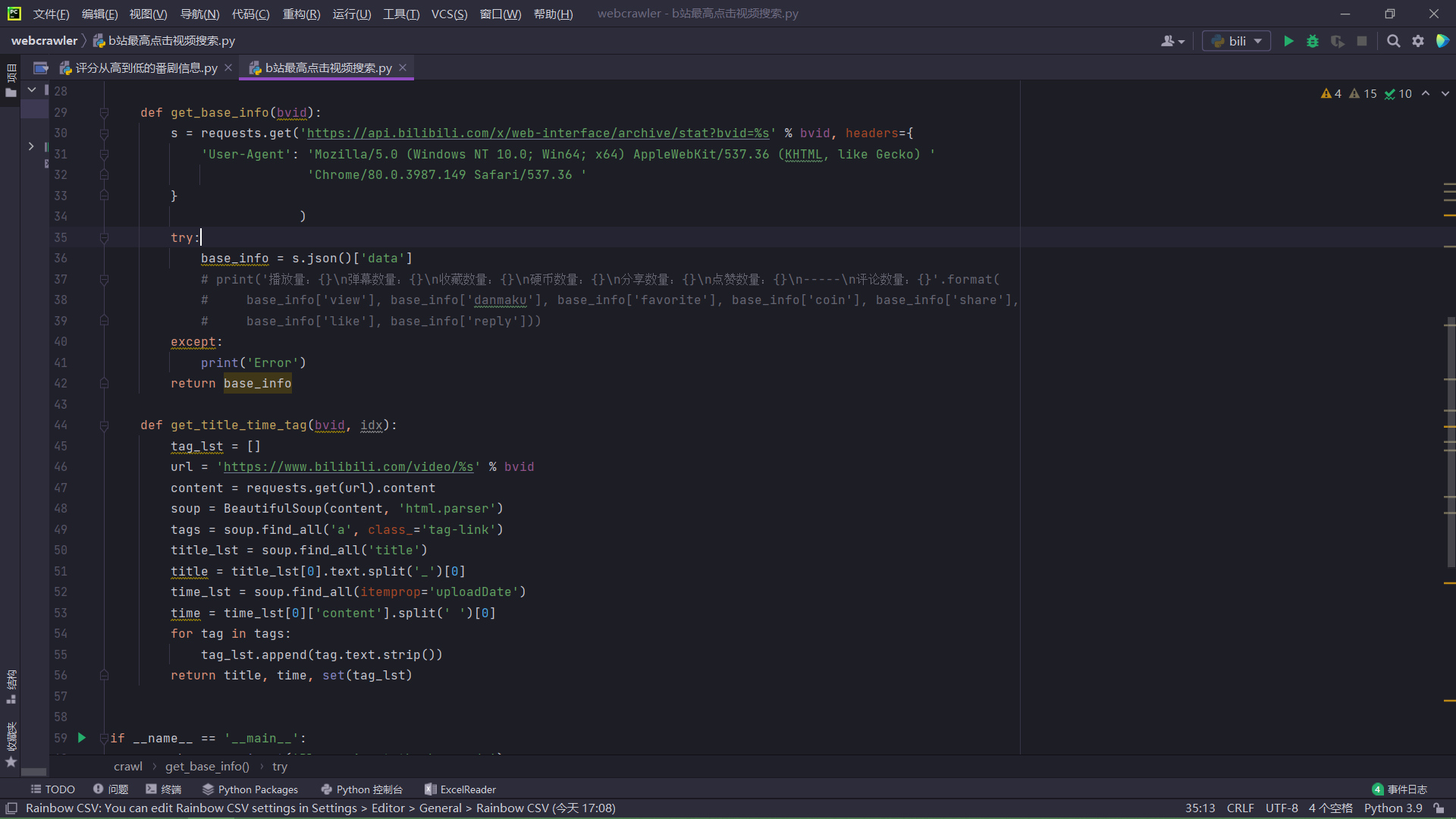
**3. Crawling example code analysis**

**3.1 search\_video**



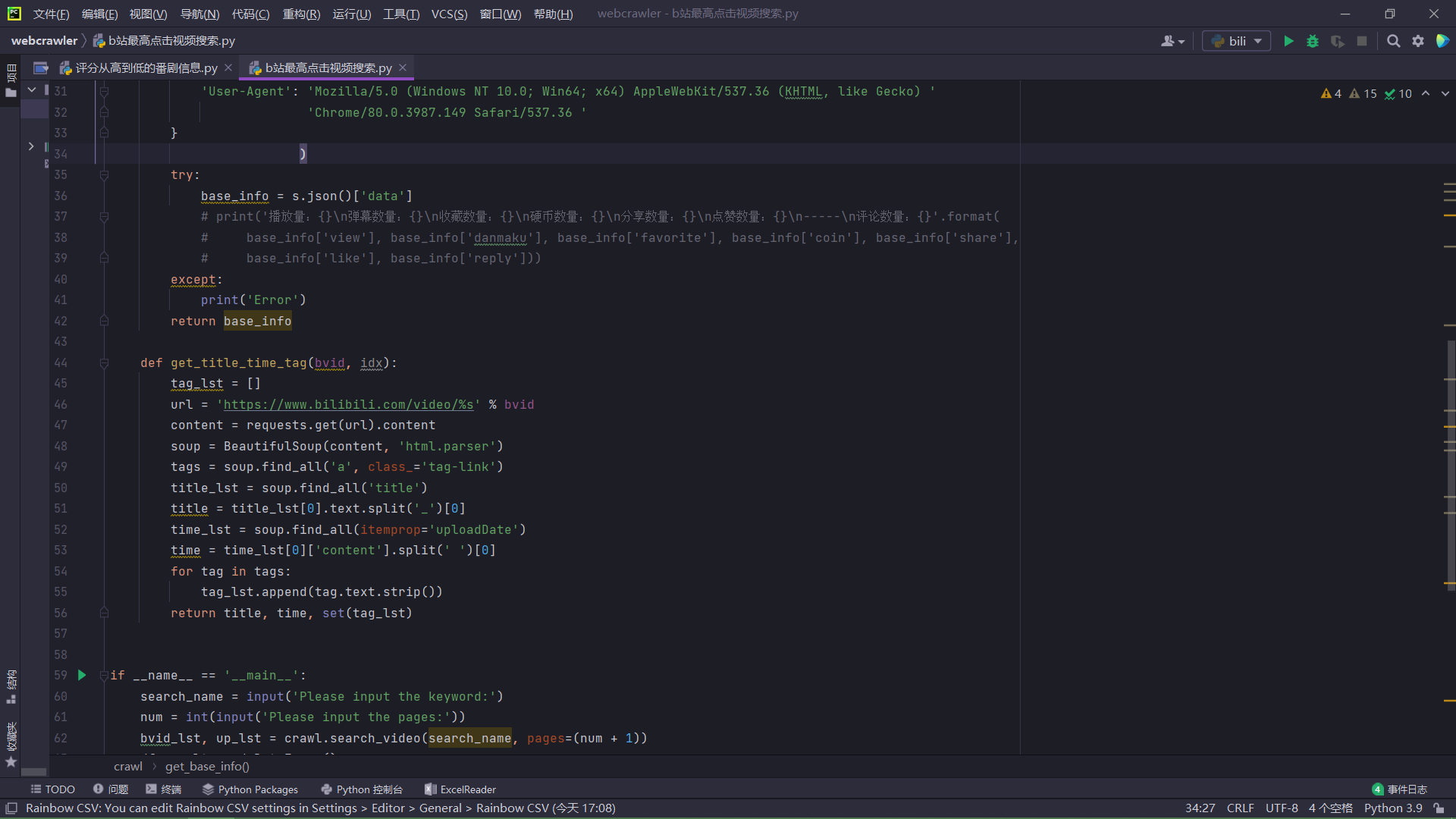
Firstly, search the videos of related topics in the search interface, and the UP master and BV number corresponding to the video were crawled

**2.2 get\_base\_info**



Analyze the JSON file on the corresponding interface and crawl the relevant data

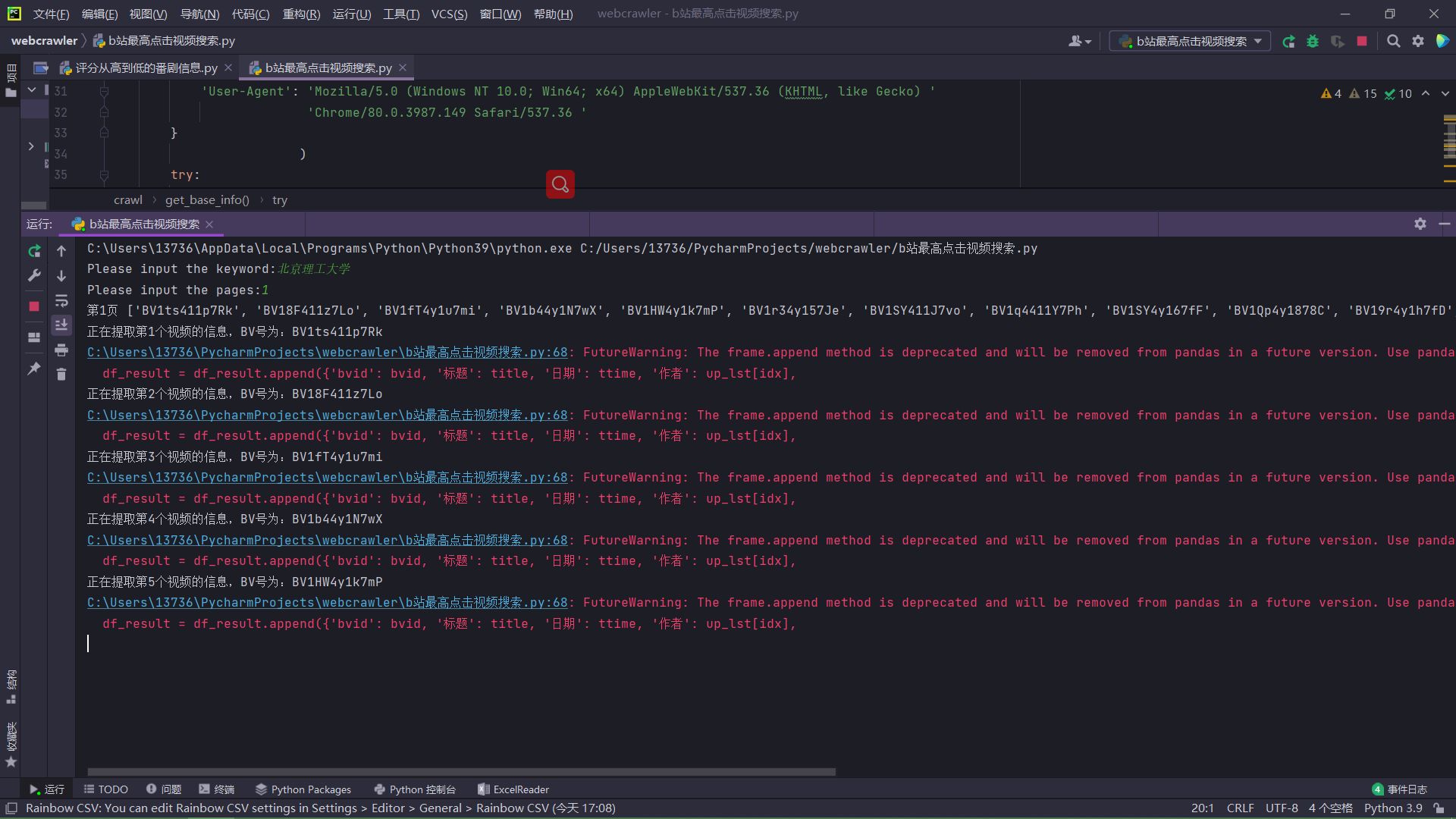
**2.3 get\_title\_time\_tag**



The HTML code of the corresponding web page is crawled, and the data of the relevant interface is found and recorded into the dictionary

**2.4 data crawling results**

The climbing process is shown in the figure:



By the way, one page consists 20 results, so if input num == 1, there will be

20 results, and it will be stored in a csv file with name

‘search\_name\_count\_num.csv’.

