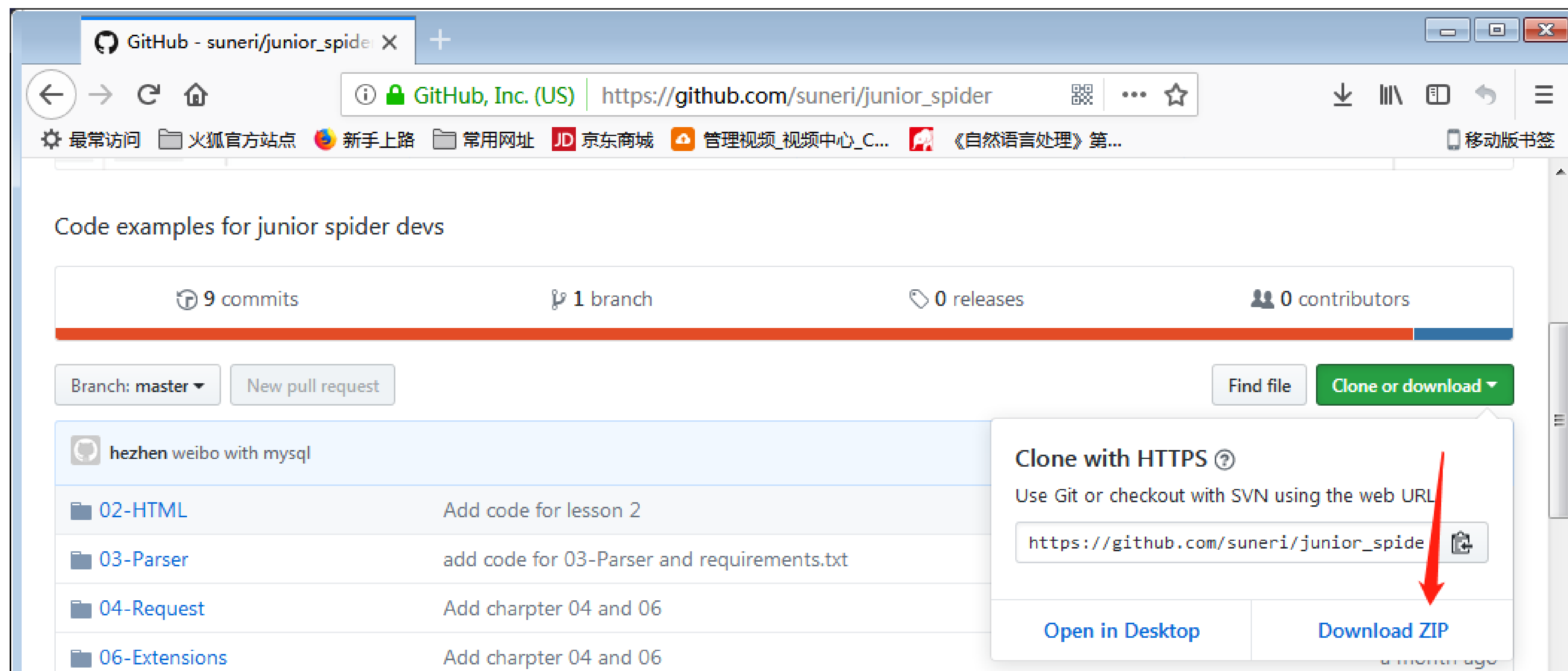


# Win下使用VSCode打开课程代码

## 主讲：雷老师

# 课程代码

- 在浏览器中打开[https://github.com/suneri/junior\\_spider](https://github.com/suneri/junior_spider)
- 点击 Download ZIP 下载课程代码



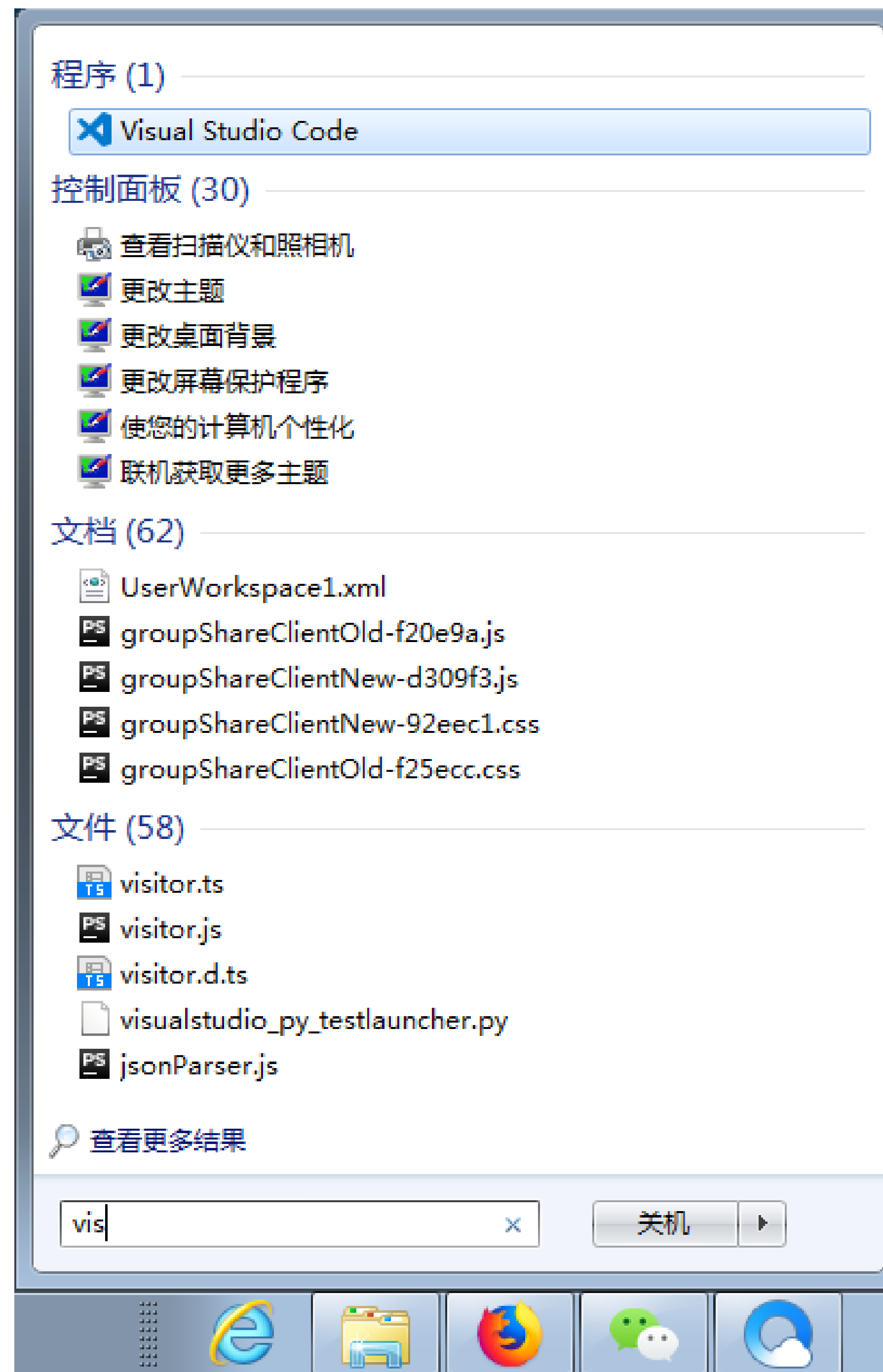
# 课程代码

- 解压junior\_spider-master.zip文件，目录显示：

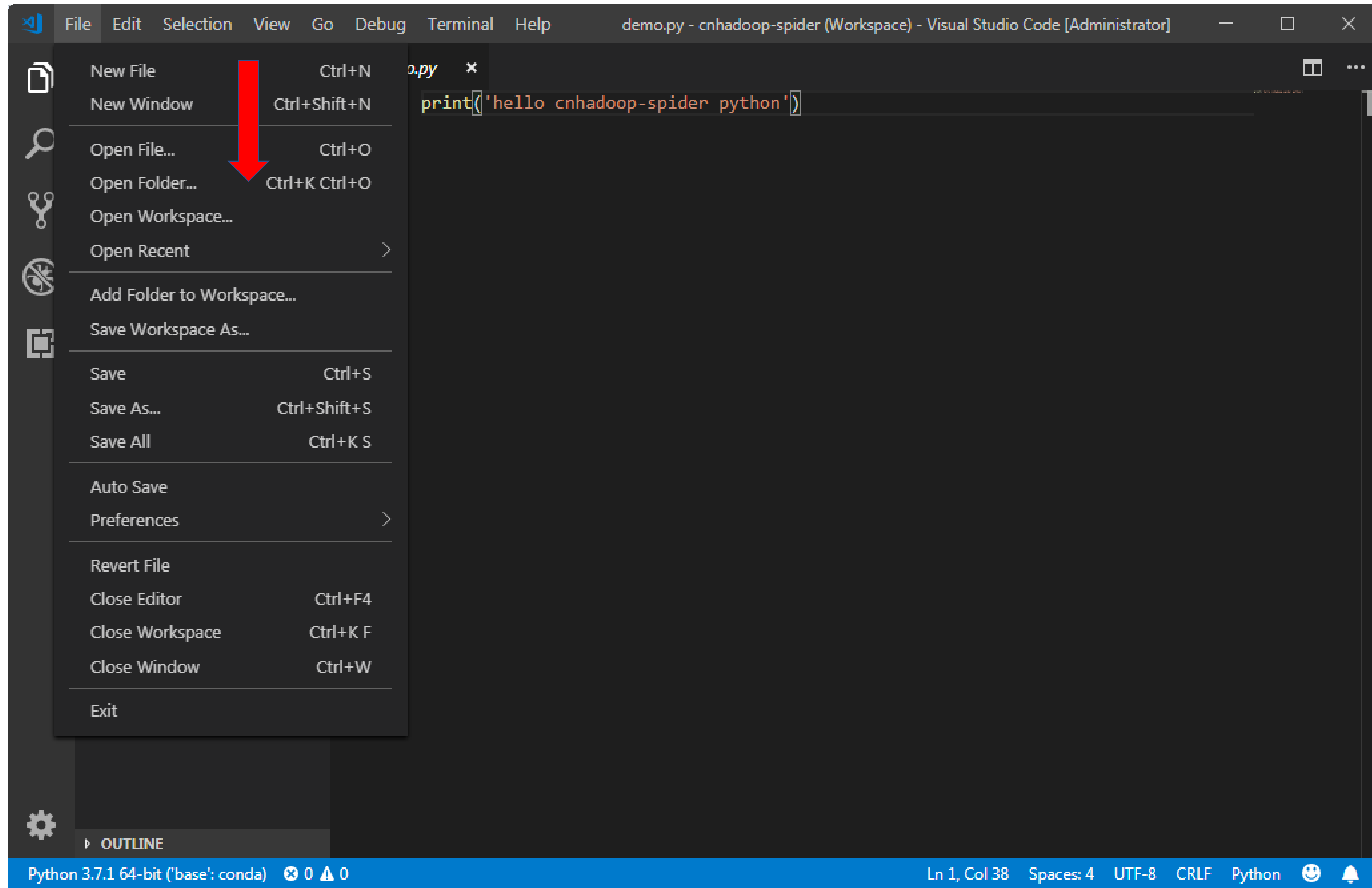
本地磁盘 (D:) ▸ DevProjects ▸ junior_spider-master ▸				
工具(T) 帮助(H)				
打印 刻录 新建文件夹				
名称	修改日期	类型	大小	
02-HTML	2019/1/16 2:44	文件夹		
03-Parser	2019/1/16 2:44	文件夹		
04-Request	2019/2/11 13:53	文件夹		
06-Extensions	2019/1/16 2:44	文件夹		
07-BBS	2019/2/11 15:04	文件夹		
09-MySQL	2019/1/16 2:44	文件夹		
10-Threads	2019/1/16 2:44	文件夹		
11-Login	2019/1/16 2:44	文件夹		
12-Weibo	2019/1/16 2:44	文件夹		
README.md	2019/1/16 2:44	MD 文件	2 KB	
requirements.txt	2019/1/16 2:44	文本文档	1 KB	

# 打开VSCode

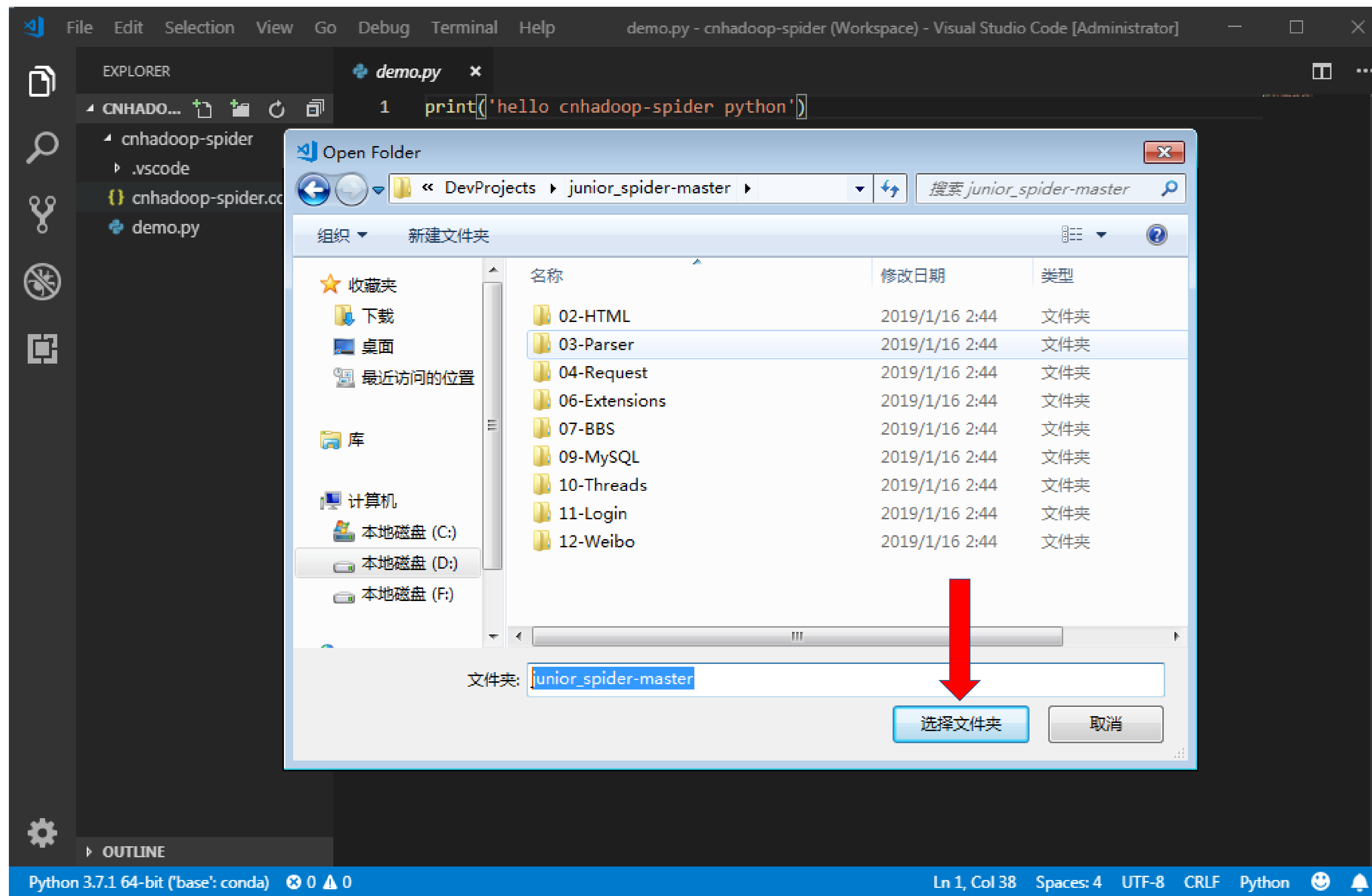
- 安装Anconada时也安装了ide 工具VSCode，在Windows下点击“开始”，输入visual 可以搜索到的Visula Studio Code程序图标。单击程序图标即可打开。



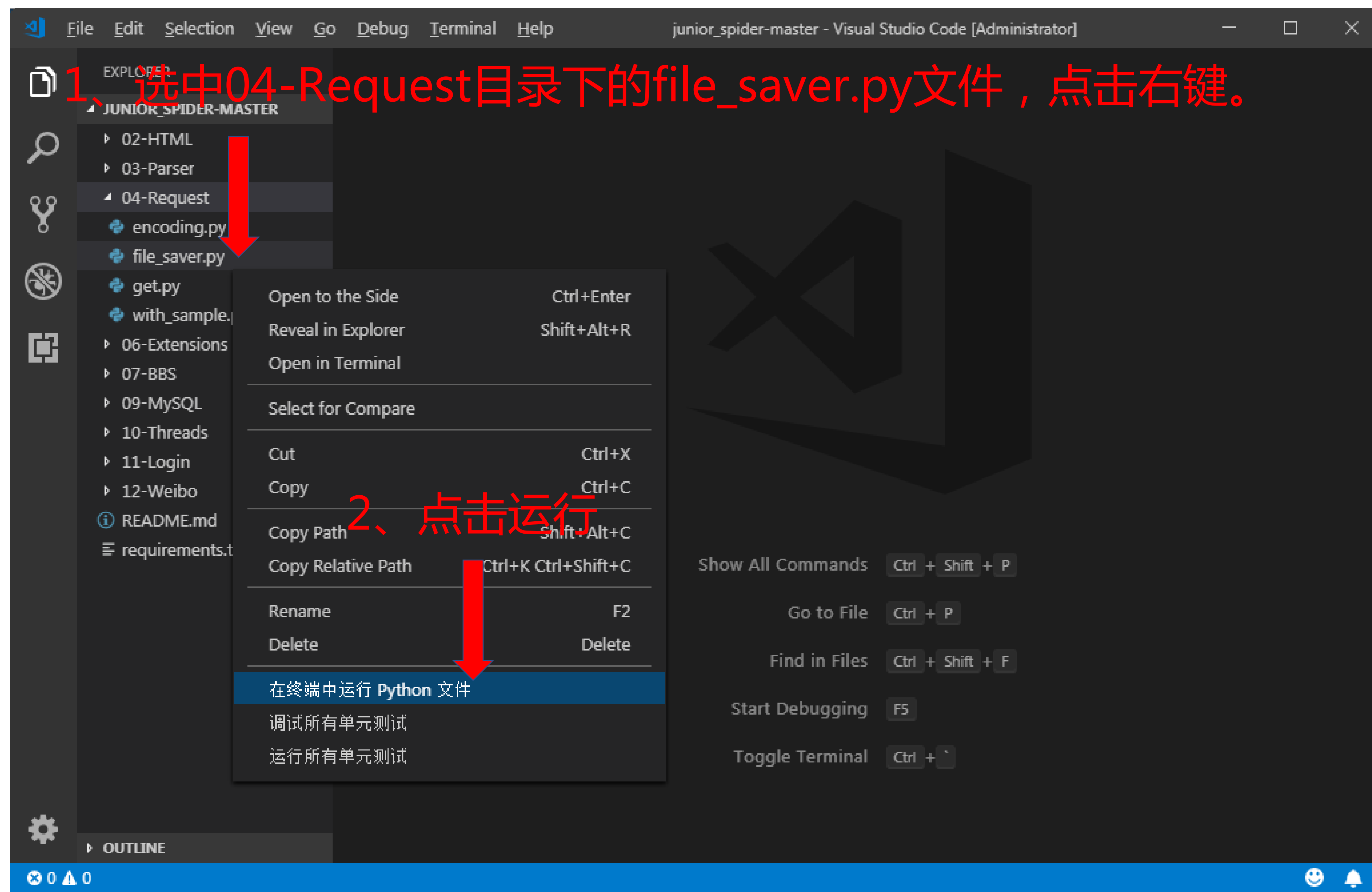
# 点击左上角File菜单项Open Folder...



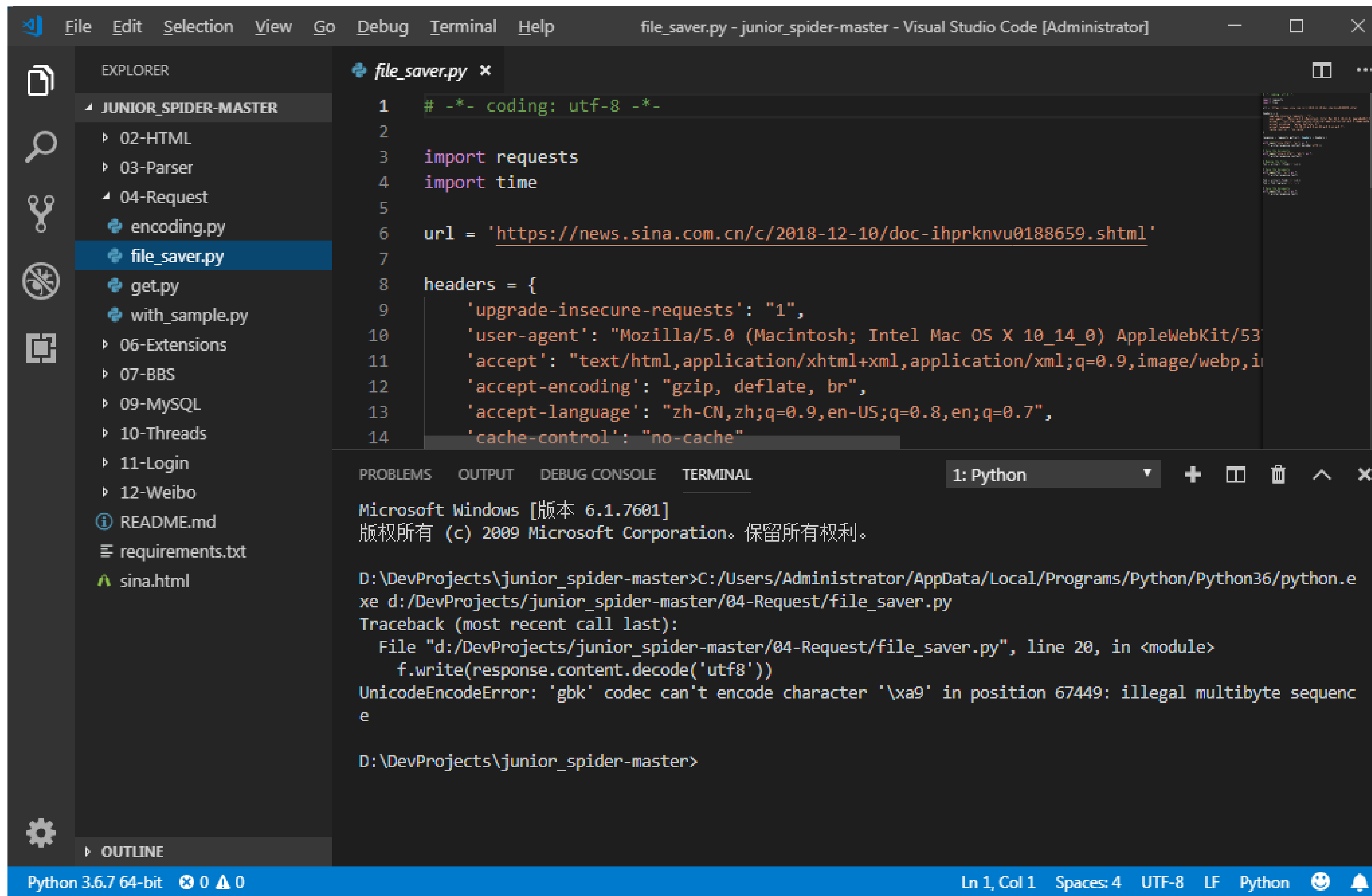
# 选中下载的课程代码目录



# 运行一个课程代码样例



# 在Win下直接执行会报gbk错误



The screenshot shows the Visual Studio Code interface with the following components:

- EXPLORER:** A file tree on the left showing the project structure. The file `file_saver.py` is selected under the `04-Request` folder.
- EDITOR:** The main workspace showing the content of `file_saver.py`. The code is as follows:

```
1  # -*- coding: utf-8 -*-
2
3  import requests
4  import time
5
6  url = 'https://news.sina.com.cn/c/2018-12-10/doc-ihprknvu0188659.shtml'
7
8  headers = {
9      'upgrade-insecure-requests': "1",
10     'user-agent': "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_0) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.102 Safari/537.36",
11     'accept': "text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8",
12     'accept-encoding': "gzip, deflate, br",
13     'accept-language': "zh-CN,zh;q=0.9,en-US;q=0.8,en;q=0.7",
14     'cache-control': "no-cache"
```
- TERMINAL:** The bottom panel showing the command prompt output. It displays the command to run the script and the resulting error:

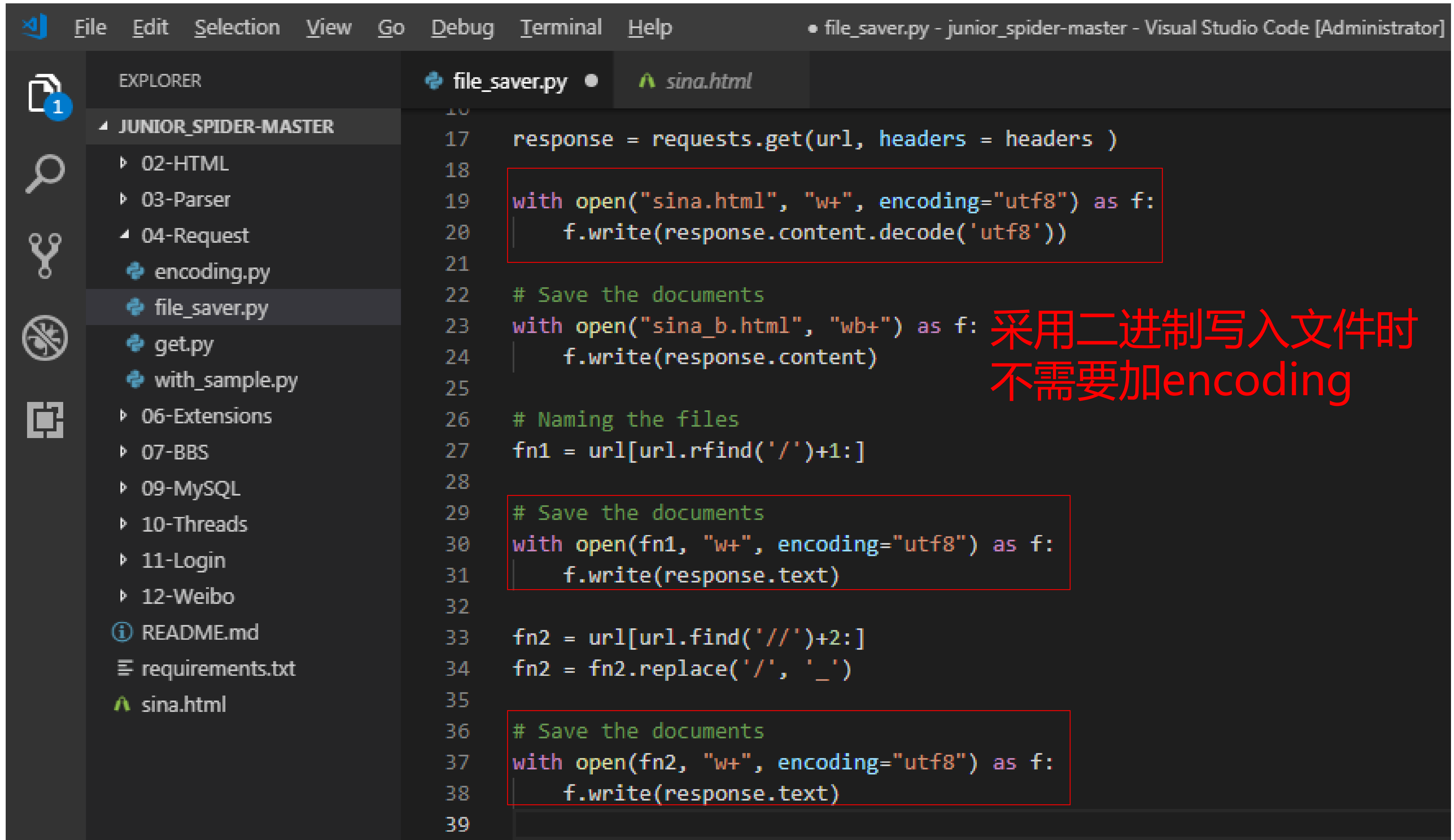
```
D:\DevProjects\junior_spider-master>C:/Users/Administrator/AppData/Local/Programs/Python/Python36/python.exe d:/DevProjects/junior_spider-master/04-Request/file_saver.py
Traceback (most recent call last):
  File "d:/DevProjects/junior_spider-master/04-Request/file_saver.py", line 20, in <module>
    f.write(response.content.decode('utf8'))
UnicodeEncodeError: 'gbk' codec can't encode character '\xa9' in position 67449: illegal multibyte sequence

D:\DevProjects\junior_spider-master>
```

The status bar at the bottom indicates the Python 3.6.7 64-bit environment is active, with 0 errors and 0 warnings.



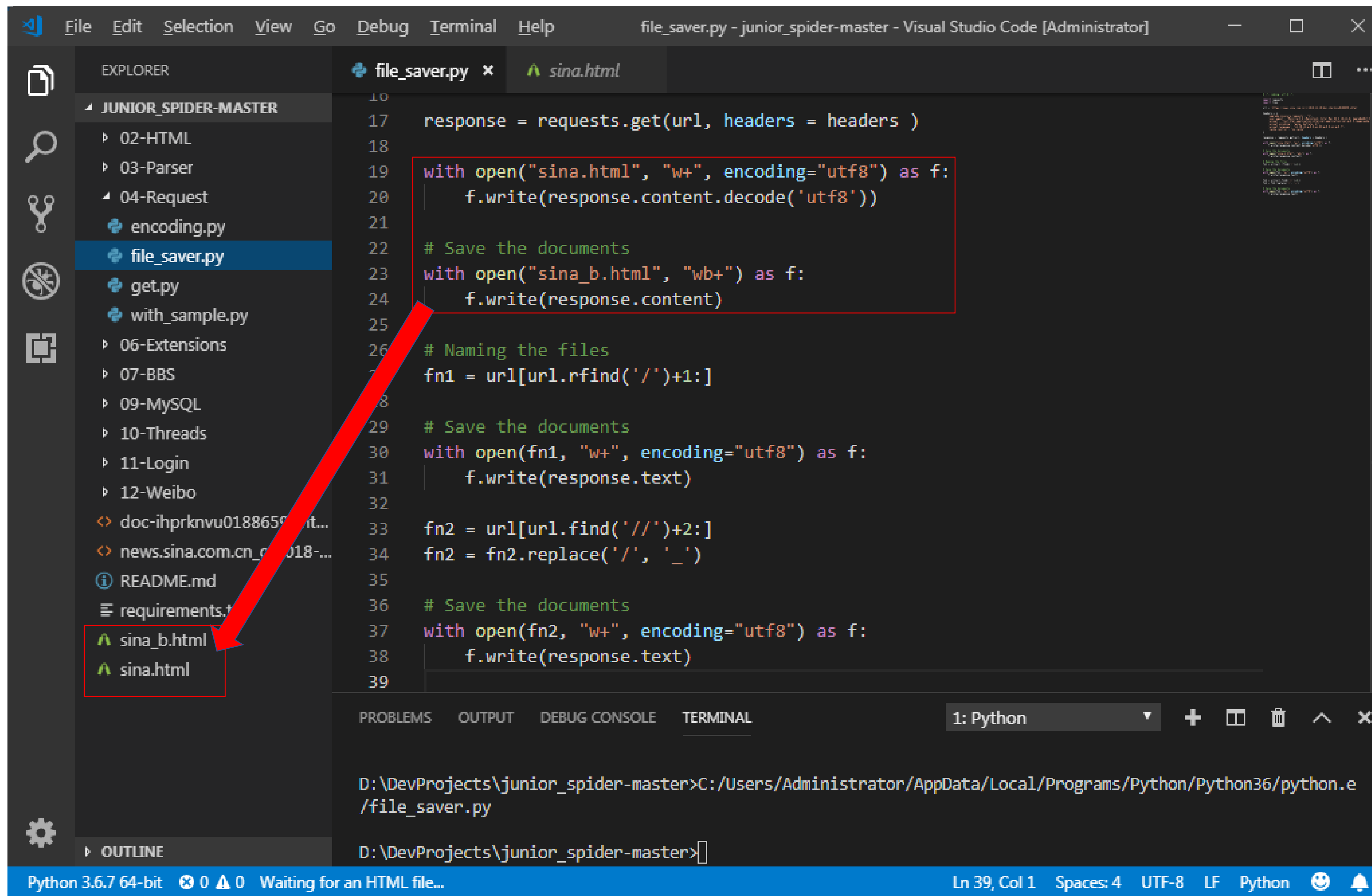
# 在with open里添加encoding="utf8"



```
10
17 response = requests.get(url, headers = headers )
18
19 with open("sina.html", "w+", encoding="utf8") as f:
20     f.write(response.content.decode('utf8'))
21
22 # Save the documents
23 with open("sina_b.html", "wb+") as f:
24     f.write(response.content)
25
26 # Naming the files
27 fn1 = url[url.rfind('/')+1:]
28
29 # Save the documents
30 with open(fn1, "w+", encoding="utf8") as f:
31     f.write(response.text)
32
33 fn2 = url[url.find('///')+2:]
34 fn2 = fn2.replace('/', '_')
35
36 # Save the documents
37 with open(fn2, "w+", encoding="utf8") as f:
38     f.write(response.text)
39
```

采用二进制写入文件时  
不需要加encoding

# 再次执行后会生成文件



```
File Edit Selection View Go Debug Terminal Help file_saver.py - junior_spider-master - Visual Studio Code [Administrator]

EXPLORER
JUNIOR_SPIDER-MASTER
├── 02-HTML
├── 03-Parser
├── 04-Request
│   ├── encoding.py
│   ├── file_saver.py
│   ├── get.py
│   └── with_sample.py
├── 06-Extensions
├── 07-BBS
├── 09-MySQL
├── 10-Threads
├── 11-Login
├── 12-Weibo
├── doc-ihprknvu0188659...
├── news.sina.com.cn_018-...
├── README.md
├── requirements.t
├── sina_b.html
└── sina.html

10
17 response = requests.get(url, headers = headers )
18
19 with open("sina.html", "w+", encoding="utf8") as f:
20     f.write(response.content.decode('utf8'))
21
22 # Save the documents
23 with open("sina_b.html", "wb+") as f:
24     f.write(response.content)
25
26 # Naming the files
27 fn1 = url[url.rfind('/')+1:]
28
29 # Save the documents
30 with open(fn1, "w+", encoding="utf8") as f:
31     f.write(response.text)
32
33 fn2 = url[url.find('///')+2:]
34 fn2 = fn2.replace('/', '_')
35
36 # Save the documents
37 with open(fn2, "w+", encoding="utf8") as f:
38     f.write(response.text)
39

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL 1: Python

D:\DevProjects\junior_spider-master>C:/Users/Administrator/AppData/Local/Programs/Python/Python36/python.e
/file_saver.py

D:\DevProjects\junior_spider-master>
```

Python 3.6.7 64-bit 0 0 Waiting for an HTML file... Ln 39, Col 1 Spaces: 4 UTF-8 LF Python

# 课程代码仅供参考

- 课程代码仅供参考，请根据讲解视频多多练习。
- 最好跟着老师敲一遍代码。

只为遇见明天更优秀的你！