

# 小小爬虫使用说明4.3

## 依赖项及安装方法

1.先准备好能运行Python的软件 如vscode （下载vscode中的python控件）或者PyCharm等

2.下载相应的pip控件：（[官网地址](#)）

windows + r --> cmd --> 输入 pip install ( ? )

? 内是要使用的

```
1 requests
2 csv
3 re
4 time
5 pandas
6 os
7 datetime
```

## 功能介绍

在运行之后，基本可以实现每60爬取官网的通知的数据（链接，名称，发布时间）

## 各部分使用方法

打开程序运行，等待运行结束，打开生成的xlsx文件查看结果。

（如果要实现每分钟自动爬取，请不要关闭程序）(๑• 3 •๑)

（请保持4.3文件夹的SDU文件夹与程序在同一目录下）

## 结果说明

生成的SDU.xlsx就是可查看的文件

A	B	C
链接	题目	日期
news.NewsContentUrl&wbtreeid=1013&wbnewsid=34942	关于组织2023届本科毕业生普通话水平测试报名工作的通知	2022-10-14

结果包括以上几列

sp?urltype=news.NewsContentUrl&wbtreeid=1015&wbnewsid=32693 | 天十编制审核2022年春季学期教学执行计划的通知  
10 19 22 26 15.csv 10 19 22 26 54.csv 10 19 22 27 33.csv 10 19 22 28 11.csv 10 19 22 28 31.csv 10 19 22 29 27.csv ..

下方是不同时间的sheet（格式为 月 日 时 分 秒）

在此结果中，没有出现https： 的均为不完整链接，请自行添加（<https://www.bkjj.sdu.edu.cn/>）部分链接

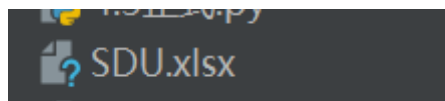
## 测试截图

```

File Edit View Navigate Code Refactor Run Tools VCS Window Help 4.3ceshi.py 4.3ceshi.py
pythonthy 4.3ceshi.py
Project pythonthy D:\pythonthy
> SDU
> view
4.1(1)正式.py
4.1.1.1.py
4.2.2.py
4.2.py
4.2 (1).py
4.2正式.py
4.3ceshi.py
4.3正式.py
SOU.xlsx
SOU通知.xlsx
学习.py
用户信息.py
相同up.csv
知字数据.csv
External Libraries
Scratches and Consoles

36 f = open(tti+str(ti)+'.csv', mode="w", newline="", encoding='gb18030')
37 csvwriter = csv.writer(f)
38 task()
39 f.close()
40 print("over!")
41 newdir = './SDU' # 列出文件夹下所有的目录与文件
42 writer = pd.ExcelWriter('SDU.xlsx')
43 list = os.listdir(newdir)
44 print(list)
45 for i in range(0, len(list)):
46     data = pd.read_csv(tti+list[i], encoding="gb18030", index_col=0)
47     data.to_excel(writer, sheet_name=list[i])
48 writer.close()
49 print("over!")
50 time.sleep(60)
51 while True
Run: 4.3ceshi
D:\python\python.exe D:\pythonthy\4.3ceshi.py
2022-10-20 13:00:24
over!
['10 20 13 00 24.csv']
over!
Find Run Version Control Python Packages TODO Python Console Problems Terminal Services
Localized PyCharm 2022.2.2 is available // Switch and restart // Don't ask again (12 minutes ago) 39/14 CRLF UTF-8 4 spaces Python 3.10

```



A	B	C	D	E	F	G	H	I	J	K	L
链接	题目	日期									
ontentUr	关于组织	2022-10-14									
ontentUr	关于做好	2022-09-16									
1. edu. cn/	第十四届	2022-09-08									
ontentUr	关于2022	2022-05-27									
ontentUr	关于做好	2022-05-26									
ontentUr	山东大学	2022-04-08									
ontentUr	山东大学	2022-03-31									
ontentUr	山东大学	2022-03-24									
ontentUr	本科生院	2022-03-24									
ontentUr	关于做好	2022-03-15									
ontentUr	山东大学	2022-03-10									
ontentUr	关于2021	2021-12-03									
ontentUr	关于做好	2021-12-02									
ontentUr	关于组织	2021-11-13									
1. edu. cn/	2021年全	2021-11-04									
链接	题目	日期									
. edu. cn/	关于在部	2022-09-30									
ontentUr	关于做好	2022-09-23									
ontentUr	关于做好	2022-09-20									
ontentUr	关于本科	2022-09-14									
ontentUr	关于2022	2022-09-07									
ontentUr	关于做好	2022-08-24									
ontentUr	关于做好	2022-08-23									
ontentUr	关于维护	2022-07-05									
ontentUr	关于2022	2022-07-01									
ontentUr	关于2022	2022-07-01									
ontentUr	关于期末	2022-06-17									
ontentUr	关于2022	2022-06-16									
ontentUr	山东大学	2022-06-14									
ontentUr	关于2022	2022-06-08									
ontentUr	关于做好	2022-06-08									
链接	题目	日期									
ontentUr	关于编制	2022-10-07									
ontentUr	关于开展	2022-09-06									
ontentUr	关于开展	2022-05-23									
ontentUr	关于开展	2022-04-21									
ontentUr	关于开展	2022-03-23									
<	>	>	10 20 13 00 24.csv	10 20 13 02 10.csv	+						