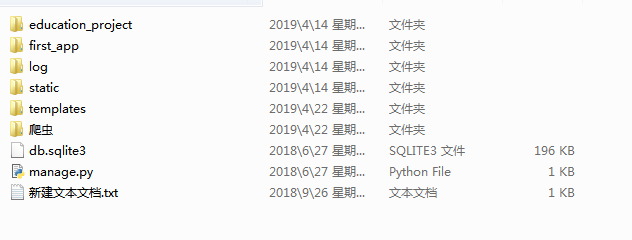
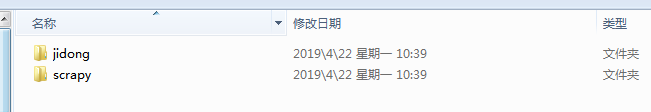
**爬虫期中作业**



pythonWeb\_venv为虚拟目录，education\_project是项目目录

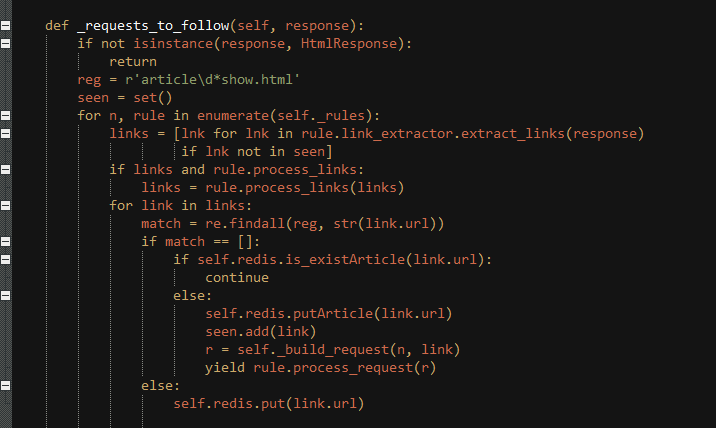


Templates 存放html，static存放资源文件，爬虫存放爬虫文件，



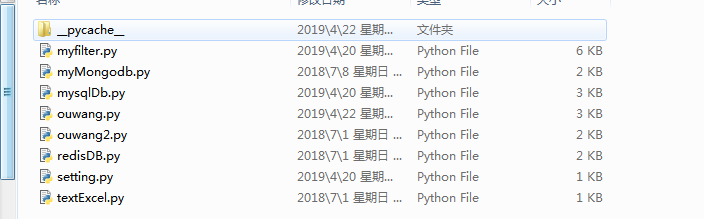
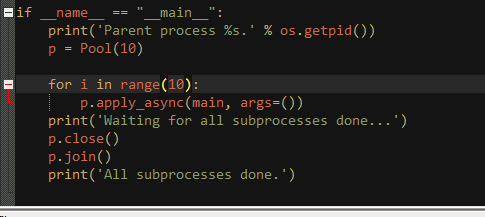
静态网站：瓯网使用scrapy爬取，

修改scrapy 的\_requests\_to\_follow方法，把url写入到redis.

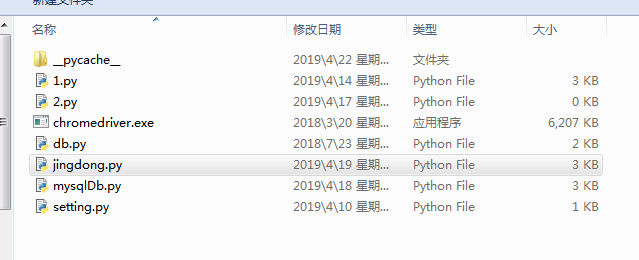


解析网页时使用多线程,使用新闻通用filter。

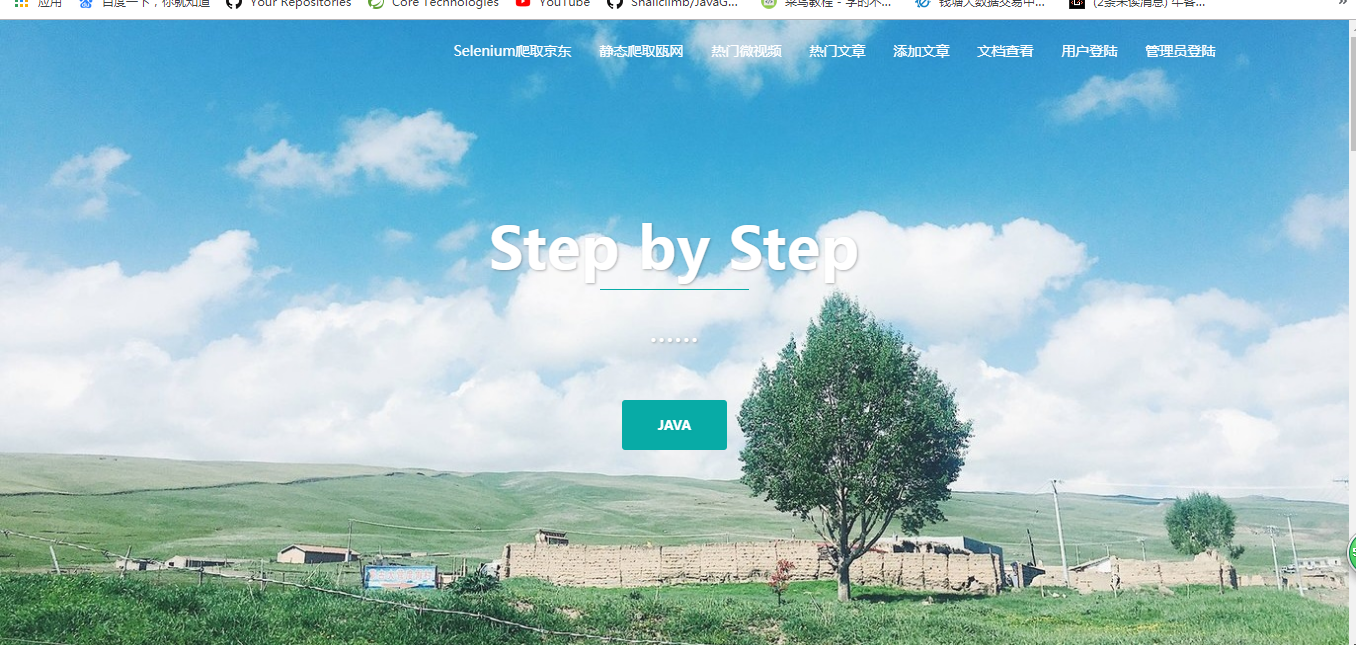




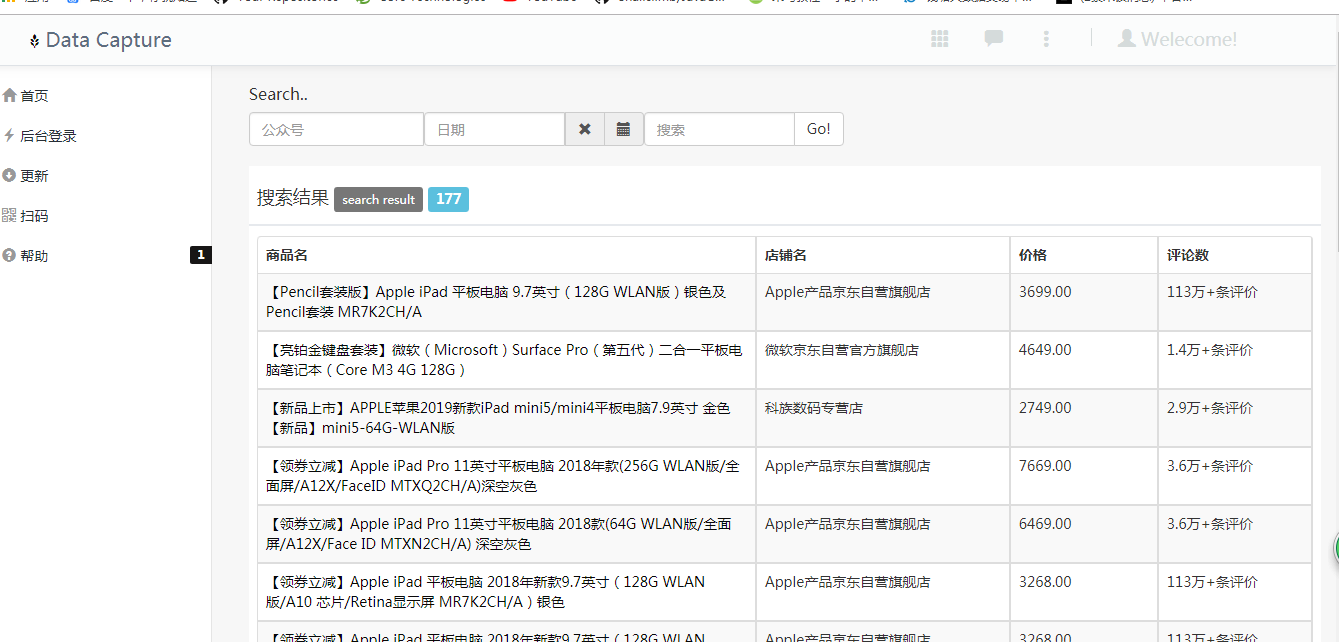
京东使用selenium爬取



**首页**



爬取京东



静态瓯网

