**功能说明**：自动下载pubmed文章脚本，默认下载前200篇，暂只支持PMC站点文章下载。仅作为学习自用，请不要用作其他途径。

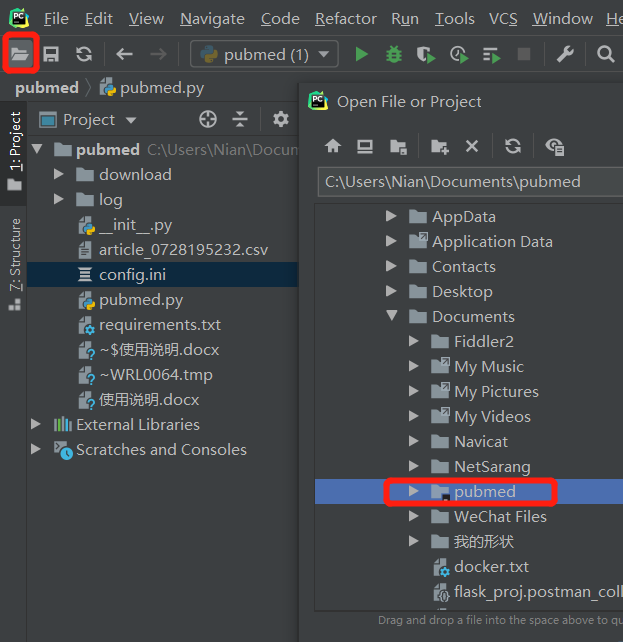
1. Python环境安装：

[python官网](http://www.python.org/)下载安装最新版本的python版本即可，python3以上都行，[安装步骤](https://www.sohu.com/a/337364638_120255642)

1. Python IDE软件：

可以安装一个pycharm社区版，不用破解，方便编写和运行程序，安装后open目录。

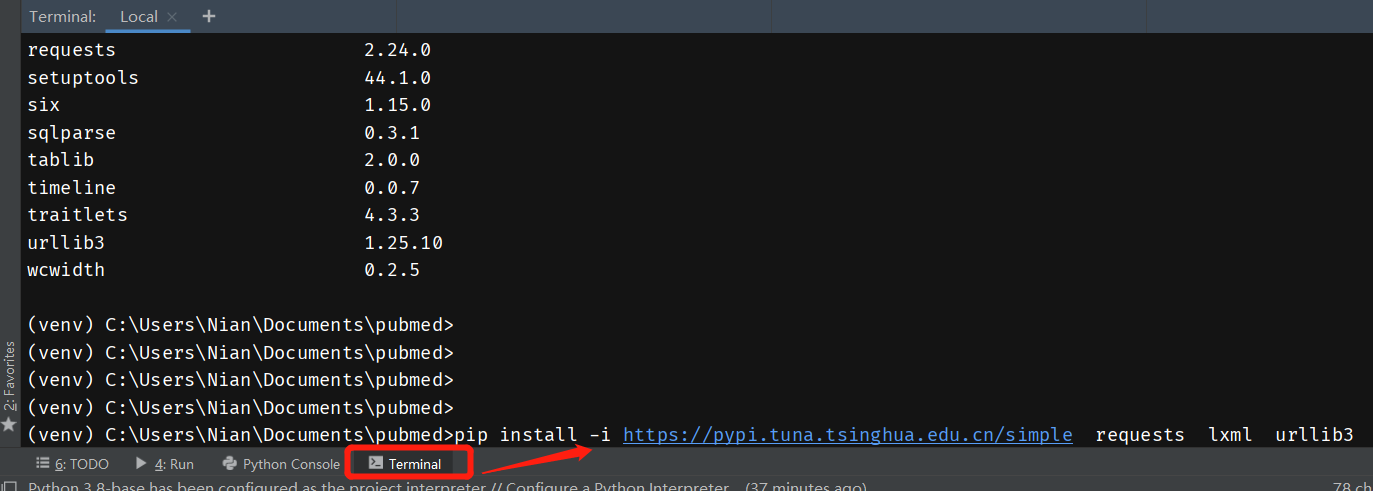
<https://download.jetbrains.8686c.com/python/pycharm-community-2020.1.4.exe>



1. 依赖的软件包安装:

pip install -i https://pypi.tuna.tsinghua.edu.cn/simple requests lxml urllib3

点击pycharm底端Terminal终端，复制上面命令后回车, (win+R后cmd终端中也可以安装)

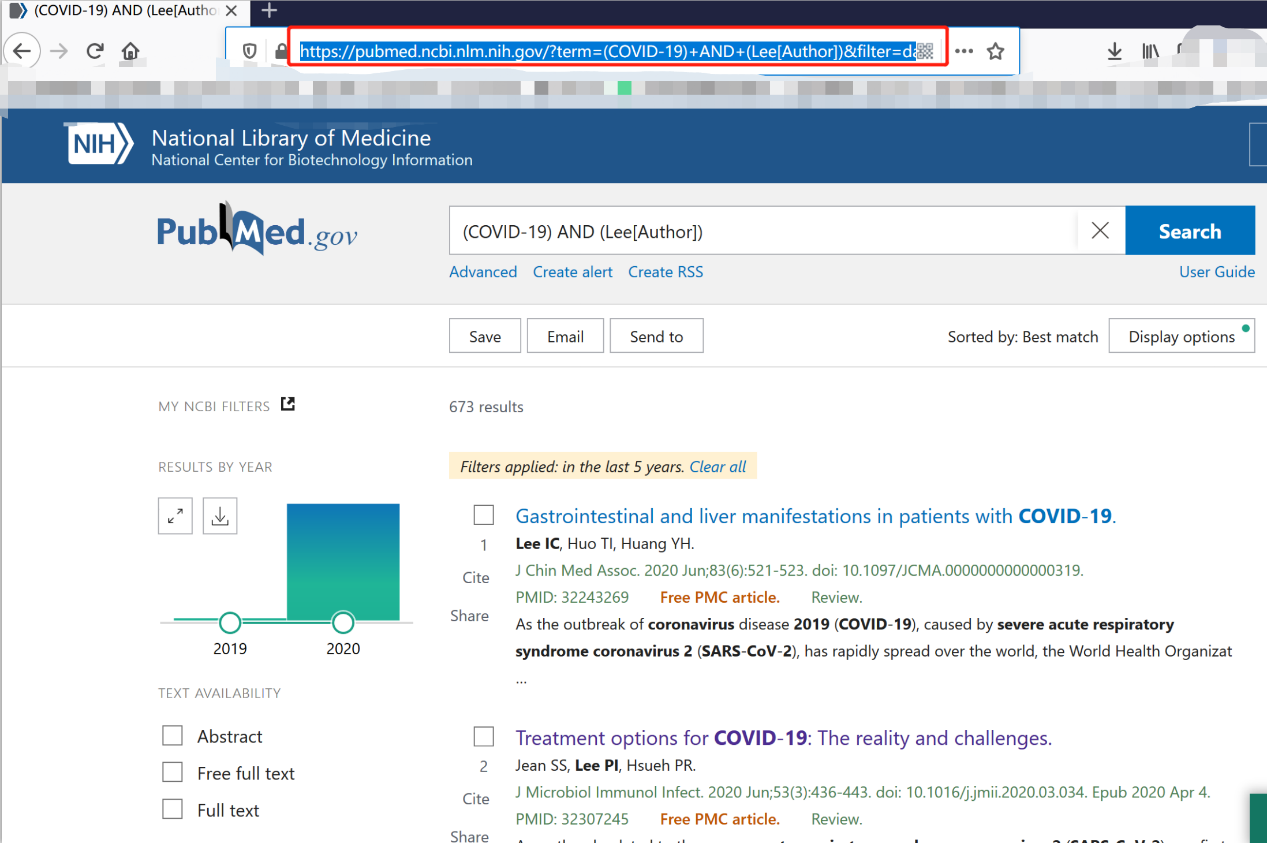


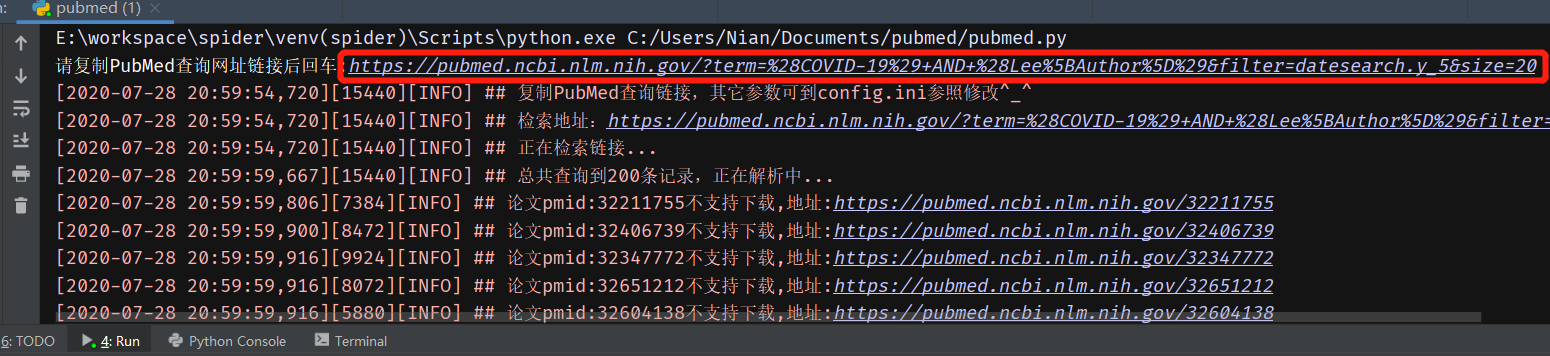
1. 运行使用: pubmed.py

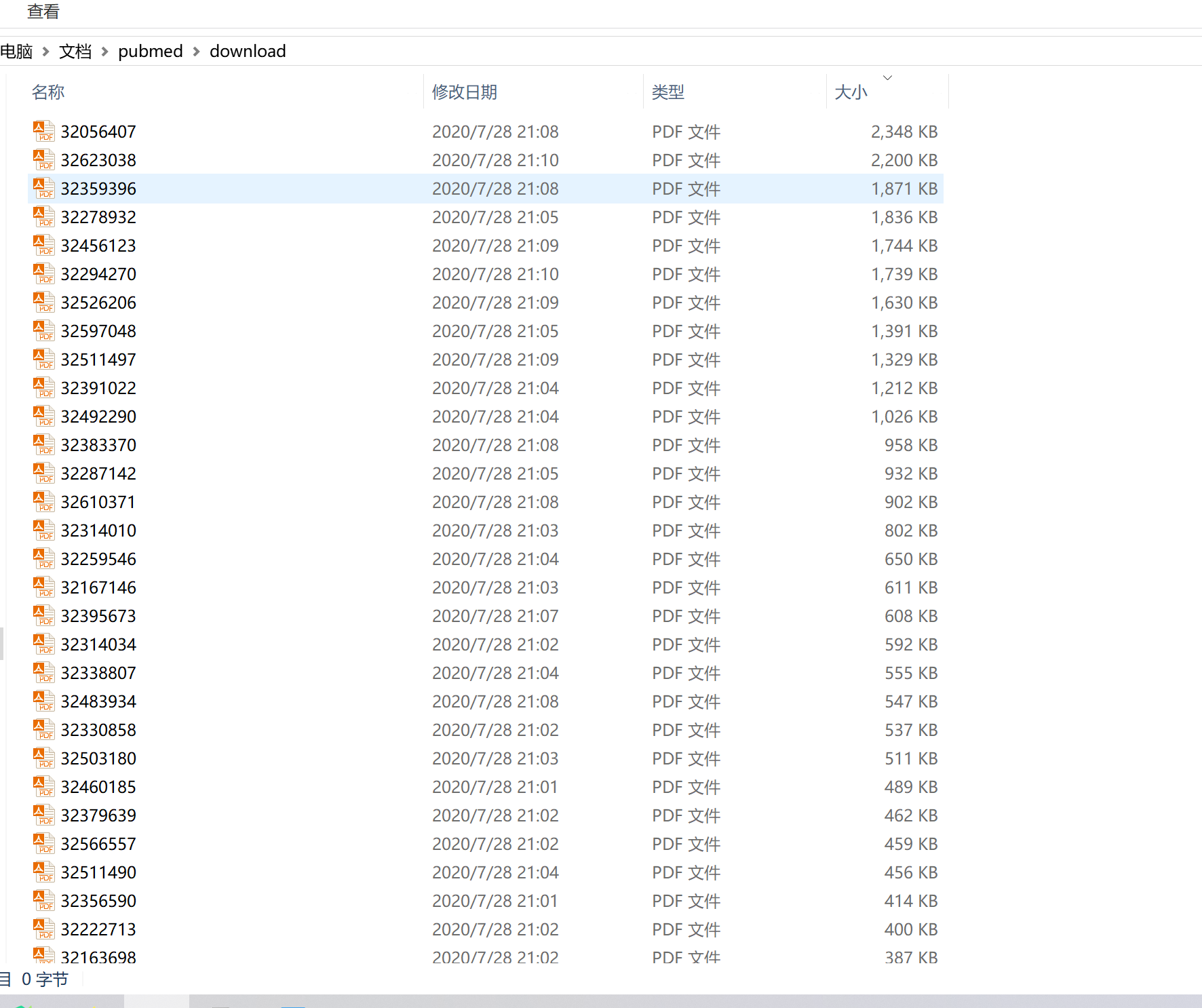
（1）打开pubmed.py脚本后，右键Run运行；

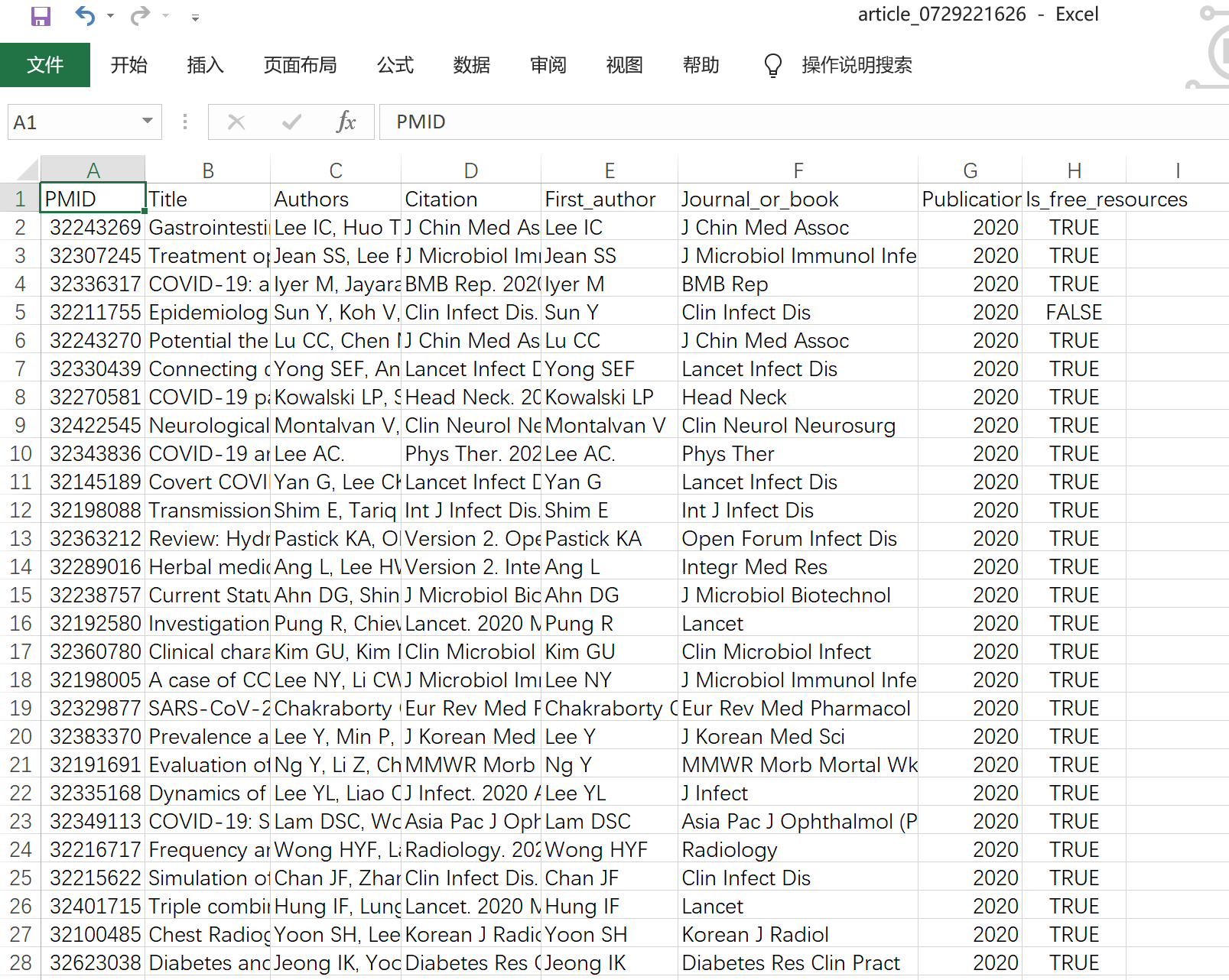
（2）正常的在PubMed网站进行普通检索或者高级检索，然后点Search，对搜索结果也可以加过滤条件，然后复制地址栏网址到运行的脚本输入框后回车，开始执行脚本，进行搜索和下载，download目录是下载的pdf（名称为论文PMID），article\_\*.csv是导出搜索结果表格，log目录是执行日志，config.ini是可以调的参数配置文件。

说明：目前只是针对PMC文章支持下载，其他站点的文章解析下载地址需匹配的情况较多，暂时没有处理下载，但是会在log里打印下载地址，可以去页面手动下载，不支持下载说明该文章不是Free article。









1. 参数修改：config.ini
2. max\_size：一次最多下载的文章数量10,20,50,100,200，暂时没做分页处理，最多200个，如果不够用，回头可以加分页处理，可支持更多；
3. thread\_num：同时并行执行的线程数，可以填20~100，太大请求速度过快，会被PubMed服务器拒绝访问，出现Connection aborted；
4. timeout：请求的超时时间，当网络不好时可以调大；
5. is\_output\_csv：是否导出搜索结果表格； yes / no
6. 其他参数可以不用管。