

关于爬虫脚本的使用说明:第一次使用时:

- 1、安装python3、最好安装谷歌浏览器Chrome v74或v75
- 2、pip3 环境变量的配置确认
- 3、pip3 install piping
- 4、进到py脚本
- 5、pipenv install
- 6、确认是否安装成功脚本所需的第三方模块，执行行pipenv graph
- 7、pipenv run python py脚本名带全路路径 result保存路路径(必须以\结尾) 需要 遍历搜索结果的页页数 浏览器驱动名带全路路径 关键词(可以多个，以空格隔 开， site:可固定搜索网网址， -()可过滤括号内关键词)

范例:

- 1、<https://www.python.org/downloads/>下载最新版python最新版安装包，本例例python3.7.3 win 64bit

在第一步中，选择python路路径加入入OS的环境变量量path，打勾后，各种默认下一步，安装完毕; 打开cmd，输入入python 确定是否已安装，如果返回信息为正常的python版本 号等信息，则确认，然后quit()

- 2、安装google chrome。目目前20190605，可下载版本为75，
- 3、cmd 中输入入 pip3，确认python环境变量量是否已配置进入入

4、cmd 中输入 `pip3 install pipenv` 安装脚本所需的虚拟环境 5、将webspider文件夹拷贝到本机某个目录，比如E:\

6、cmd 中输入 `cd E:\webspider`

7、cmd输入 `pipenv install` 以安装环境下所需的第三方库

8、cmd输入 `pipenv graph` 以确认所需库已安装成功 (此脚本需要的第三方库有:beautifulsoup4, lxml, selenium)

9、cmd 中输入 `cd E:\webspider` 进入脚本所在文件夹

10、cmd 中输入 `pipenv run python E:\webspider\spider.py E:\webspider\ E:`

`\webspider\chromedriver-Windows 3 PCS9000系统 site:wenku.baidu.com` 11、到

`E:\webspider\`目录下寻找，打开PCS9000系统\_site.csv文档，即可浏览 百度搜索前三页的结果

注意:

1、一台机器上，step1-step8只要做一遍，安装成功后，以后每次只需要在进到py脚本所在路径下，执行step9-10即可。

2、目前提供的E:\webspider\chromedriver-Windows驱动可能只支持72-75 版本的chrome浏览器，如果浏览器版本号继续升级，可能需要到<https://selenium-python.readthedocs.io/installation.html#drivers> 下载对应链接。由于谷歌浏览器官方驱动在google.com 域名下，因此下载时可能需要打开VPN。非官方免翻墙驱动下载地址:<http://chromedriver.storage.googleapis.com/index.html>

3、由于百度引擎会在到达特定页页数后自自动折叠某些重复的结果，因此不不建议 搜索超过十十页页的数据，以免造成数据冗余或异常。

4、由于百度搜索引擎限制，需要在不同指定网站下搜索关键词时，必须重新启动此程序。不支持一次搜多个指定网站下的结果。