# AJK使用操作流程

|  |  |  |  |
| --- | --- | --- | --- |
| 版本 | 作者 | 版本描述 | 完成日期 |
| 1.1 | S | 基本实现对安居客网站楼盘数据抓取及去重 | 2017.3.24 |

## 使用介绍

1.修改连接数据库文件dianping \dbconnect.py，连接到指定数据库

2. 修改项目pipelines.py文件第49，50行和第98行的数据库tables名字，以防爬虫运行过程中将已有数据表冲掉。

3.先在本地开启一个代理访问<http://gatherproxy.com/proxylist/anonymity/?t=Elite>这个网站，因为这个网站需要翻墙。再在命令行中启动代理proxy爬取进程ajk\utils\crawl-proxy.py，这样会把成功率高，高匿名的IP爬取到validProxy.txt文件中（高匿名的IP不会发现本地爬取的IP）。

4.重开一个cmd正式启动爬虫进程，并同时加上日志scrapy crawl dianping –s LOG\_FILE=xxx.txt