操

作

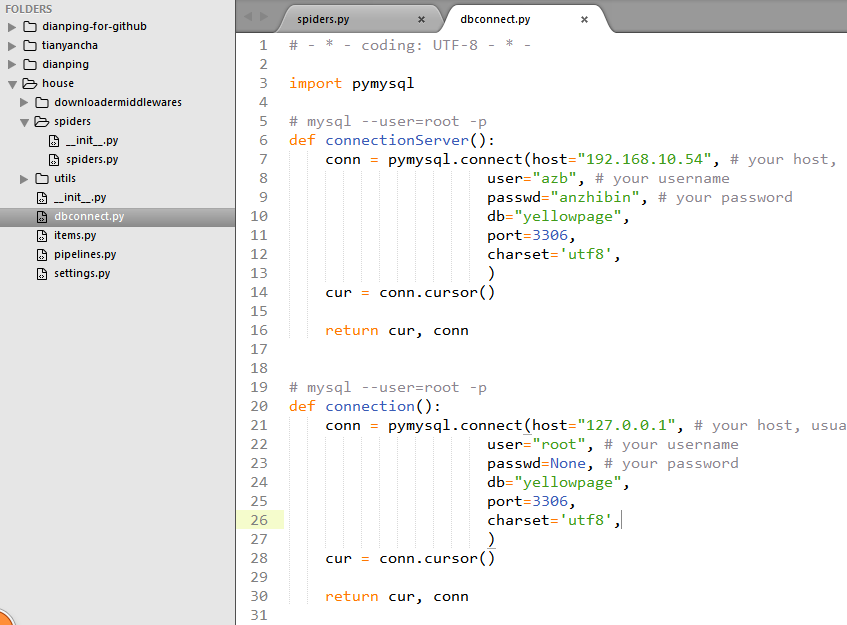
流

程

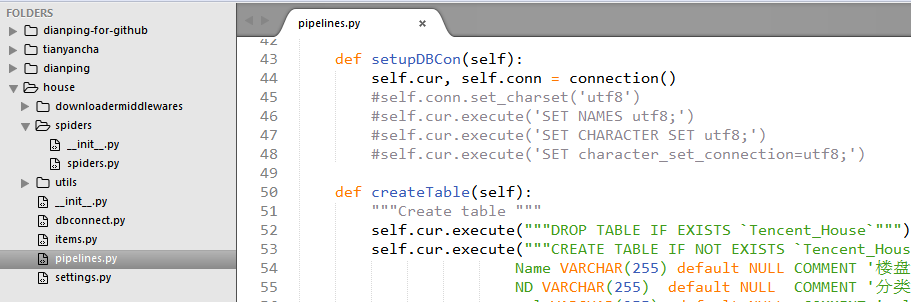
|  |  |  |  |
| --- | --- | --- | --- |
| 版本 | 作者 | 版本描述 | 完成日期 |
| 1.1 | Jerry | 基本实现对腾讯房产数据抓取及去重 | 2017.5.10 |

使用介绍

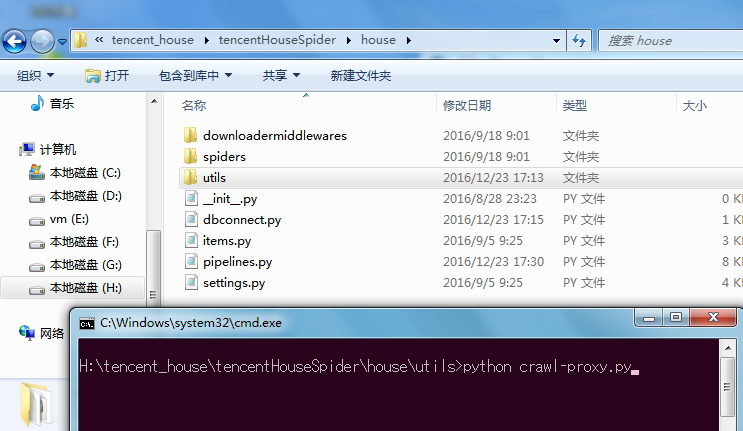
1. 修改连接数据库文件tencentHouseSpider\house\dbconnect.py，连接指定数据库



2. 修改项目tencentHouseSpider\house\pipelines.py文件第52,53行和第113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.命令行中启动代理proxy爬取进程\utils\crawl-proxy.py



4.重开一个cmd正式启动爬虫进程，并同时加上日志

