操

作

流

程

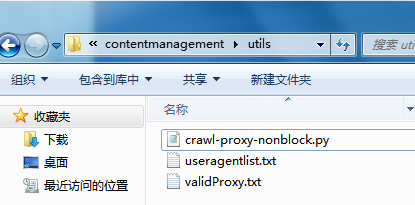
|  |  |  |  |
| --- | --- | --- | --- |
| 版本 | 作者 | 版本描述 | 完成日期 |
| 1.0 | Jerry | 基本实现对星魂黄页数据抓取及去重 | 2016.12.26 |

使用介绍

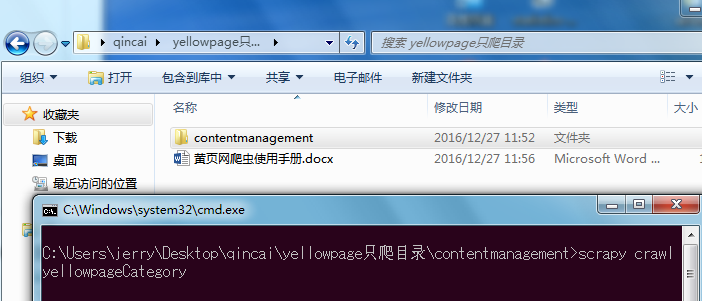
1.安装数据库browser文件



2.命令行中启动代理proxy爬取进程\utils\crawl-proxy-nonblock.py



3.重开一个cmd正式启动爬虫进程scrapy crawl yellowpageCategory，并同时加上日志



4.用sqlitebroser打开文件夹下的数据库文件Contentmanagement.db，此文件即为爬取到的星魂黄页公司一级二级三级分类数据。