**链家二手房**

**爬虫结构[ 类名 | 爬虫名 | 类所在模块]**

* **LianjiaUsedSpider** | used\_list |*/lianjia\_ershoufang/spiders/lianjia\_used.py*
* **LianjiaUsedDetailsSpider** | used\_details |*/lianjia\_ershoufang/spiders/lianjia\_used.py*

**爬虫说明**

* used\_list ：获取房源链接，将获取到的链接推倒Redis中。
  + 爬虫参数：[ city : 爬取城市名称简写]
  + 是否分布式：否
* scrapy crawl used\_list  -a city=bj/tj/sz  爬虫中， 参数city 使用 '/' 作为分隔符，城市名称列表见**其他**
* used\_details： 根据Redis 中的链接，获取房源详细信息，存储到数据库中。获得房源数据 保存到 used\_house 表中， 获取到图片链接信息 保存到 up\_img\_index 表中。
  + 爬虫参数：无
  + 是否分布式：是
* scrapy crawl used\_details  根据需求多开即可，前提是先开 used\_lis 爬虫获取链接

**数据流**

1、used\_list 获取房源链接 > 存入Redis 中。

2、used\_detail 从Redis 中获取房源链接 > 爬取房源，入库。

**其他**

|  |  |
| --- | --- |
| **城市缩写** | **城市名** |
| zs | 中山 |
| bj | 北京 |
| cd | 成都 |
| cq | 重庆 |
| cs | 长沙 |
| dg | 东莞 |
| dl | 大连 |
| fs | 佛山 |
| gz | 广州 |
| hf | 合肥 |
| hui | 惠州 |
| hz | 杭州 |
| jn | 济南 |
| lf | 廊坊 |
| nj | 南京 |
| qd | 青岛 |
| sh | 上海 |
| sjz | 石家庄 |
| su | 苏州 |
| sy | 沈阳 |
| sz | 深圳 |
| tj | 天津 |
| wh | 武汉 |
| wx | 无锡 |
| xa | 西安 |
| xm | 厦门 |
| yt | 烟台 |
| zh | 珠海 |
| zz | 郑州 |