广州链家租房信息

笔记本:	Markdown笔记		
创建时间:	2018/11/10 9:02	**更新时间:**	2018/11/10 12:53
作者:	FWJ		
URL:	https://list.yinxiang.com/markdown/eef42447-db3f-48ee-827b-1bb34c03eb83.php		

环境

- 系统：deepin 15.7
- ide：vs code
- python：Anaconda 3

fake_useragent库

随机生成请求头
安装：

```
pip install fake-useragent
```

参考资料：
https://www.jianshu.com/p/a92c4b80cc71

beautifulsoup库

官方文档：
https://www.crummy.com/software/BeautifulSoup/bs4/doc/index.zh.html#id5

解析器使用python标准库

方法主要使用find和find_all

字符串方法

切片：split()
参考资料：
http://www.runoob.com/python/att-string-split.html
去除首尾不必要字符：strip()
参考资料：
http://www.runoob.com/python/att-string-strip.html

解析器：
Python标准库
使用方法：
BeautifulSoup(markup, "html.parser")
优势：
Python的内置标准库
执行速度适中
文档容错能力强
劣势：
Python 2.7.3 or 3.2.2)前的版本中文档容错能力差

主要爬取信息

例子：
网址：
https://gz.lianjia.com/zufang/108400025954.html

数据保存

利用pandas.to_csv()

问题

1. python异常处理
   参考资料：
   https://www.cnblogs.com/klchang/p/4635040.html
   http://www.runoob.com/python/python-exceptions.html