# 我的爬虫V0.1

## 版本需求：

我的爬虫为抓取数据的入手练习，后续如有更改或增加需求在本文档后另行标注，并更新迭代版本。V1.0版本主要体现在以下几个需求：

1. 以树莓派为搭载平台；
2. 数据量应尽量节减高效；
3. 抓取数据源为房天下网站（<http://yz.fang.com/>），以获取全国、省、市按月、周为单位的房价数据，获取数据主要以热门楼盘、二手房、均价及其同比环比增减为基础，并将数据按时间、类型、区域为标签存入数据库；
4. 抓取时间按天为单位，以周为周期实现自动抓取，将每周抓取的数据做简单统计分析后以电子邮件方式发送至邮箱，以供查阅。
5. 设计敏感标记，以电子邮件返回发送方式读取，对敏感标记的数据进行跟踪，按设定方式实时采集、分析并发送数据至邮箱。

## 抓取数据类型

1. 城市：北京 上海 杭州 南京 苏州 扬州 太原 忻州

<http://newhouse.yz.fang.com/house/baojia.htm>

<http://fangjia.fang.com/yz/>

http://newhouse.yz.fang.com/house/saledate/%b1%be%d4%c2.htm?ctm=1.yz.xf\_search.head.135

<http://esf.yz.fang.com/>

http://newhouse.yz.fang.com/house/asp/trans/cjnewhouse/default.htm