**爬虫组数据采集需求**

**编制:雷邓君**

**项目名称:公平价二手车估值系统升级**

**本文件修订历史信息:**

|  |  |  |  |
| --- | --- | --- | --- |
| **更新日期** | **修订者** | **批准人** | **主要修订摘要** |
| 2018.11.05 | 雷邓君 | 万可文 | 新编写文件 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 1.需求概况

## 1.1目标

公平价二手车估值系统升级项目需要爬虫组抓取相关平台车型库，平台车源数据等信息作为估值系统算法的数据支撑。

# 2.基础数据需求

根据指定的数据来源,交易类型,车况,抓取竞争对手与我方对应的二手车估值的中间值.

## 2.1汽车之家车型库

需要抓取汽车之家所有在售/已售的汽车车型，以下为建议抓取路径和存储表。



每个款型都需要抓取对应的配置，如下



存储表:china\_used\_car\_estimate.base\_origin\_crawl\_car\_detail(sql:show create table china\_used\_car\_estimate. base\_origin\_crawl\_car\_detail 查看注释)

## 2.2 经销商/4S店新车促销价

需要抓取汽车之家所有在售新车各省的经销商/4S店促销价，地址和电话等信息，以下为建议抓取路径和存储表。

存储表:china\_used\_car\_estimate.base\_origin\_crawl\_new\_car\_4S\_price (sql:show create table china\_used\_car\_estimate.base\_origin\_crawl\_new\_car\_4S\_price 查看注释)

## 2.3 瓜子二手车平台抓取升级

注意类似”准新车”等小标题也要抓取







## 2.4 人人车二手车平台抓取升级

注意”超值”“急售“等小标题也要抓取







判断是否有以下图片(有以下图片说明价格合理):

