**汽车领域词典生成**

1. **处理流程**

从太平洋汽车网，汽车之家以及车主之家网站爬取的初始数据集，经过统计得到一定的规则，利用这些规则来处理生成汽车领域词典。

处理仅供参考，最终结果为final\_user.txt

1. **类、函数介绍**

Chezhuzhijia\_extract.py, taipingyangcars\_extract.py, qichezhijia\_extract.py分别处理不同的数据

函数的Path表示输入文件的路径。

* chezhuzhijia\_extract.py：处理车主之家数据
* deal\_chezhuzhijia (file\_path,result\_path):

输入：file\_path：输入数据文件名，chezhuzhijiacars.json

输出 ：result\_path: 输出数据文件名

所有的数据均存储在自定义的path路径下。

* taipingyangcars\_extract.py：处理太平洋汽车网的数据
  + deal\_taipingyangcars(file\_path,result\_path)

输入：file\_path：输入数据文件名，taipingyangcars.json

输出 ：result\_path: 输出数据文件名，extract

所有的数据均存储在自定义的path路径下。

* qichezhijia\_extract.py：处理汽车之家数据,得到汽车之家处理后的数据在extract\_qichezhijia.txt中
* 三个网站结果，集合，去重，过滤, 得到最终纯净版视角final\_user.txt