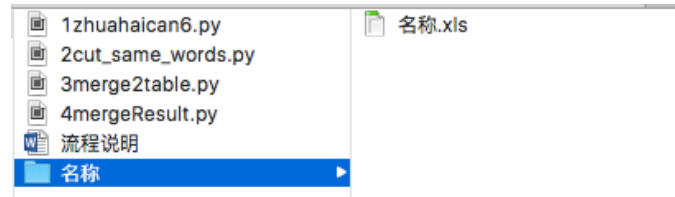


1. 在程序同一目录下新建“名称”文件夹，并将企查查下载的数据“名称.xls”放入其中



2. 实现从高德上爬取数据

- a) 在 1zhuahaican6.py 中修改一处名称

```
filename='名称'  
df_2=pd.read_excel(filename+'.xls',encoding='utf8')
```

- b) 然后修改 key,

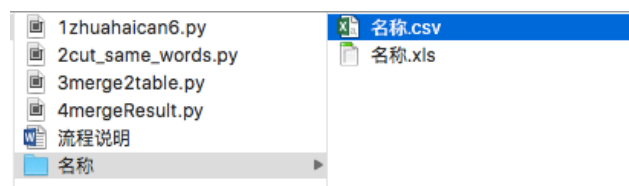
```
.....  
amap_web_key = '0864776b66cb3ca0f259a2a737bce24d'  
if i%50==0:  
    print(i)  
if i>=2000:  
    amap_web_key = '0d0f94348a01cedc7d1c96e9f871894f'  
if i>=4000:  
    amap_web_key = 'dee4a77852137fdd345d83fa480f7fd1'
```

- c) 修改之后查看第一个 key 的配额，如果第一个 key 还剩下 500 次可查，则可修改程

```
amap_web_key = '0864776b66cb3ca0f259a2a737bce24d'  
if i%50==0:  
    print(i)  
if i>=500:  
    amap_web_key = '0d0f94348a01cedc7d1c96e9f871894f'  
if i>=2500:  
    amap_web_key = 'dee4a77852137fdd345d83fa480f7fd1'  
if i>=4500:  
    amap_web_key = '137fdd345d83fa480f7fd1dee4a77852'
```

序如下

- d) 然后运行程序，输出文件“名称.csv



3. 删除两个表中企业名称中“有限公司”等通用文字

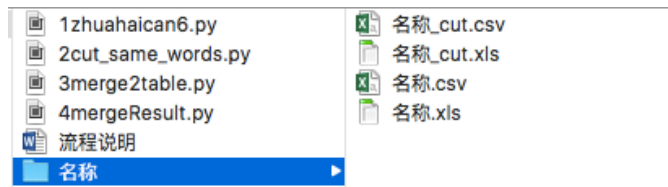
- a) 在 2cut\_same\_words.py 中修改一处名称

```
filename='名称'
typename='.csv'
inputpath=filename+'/'+filename+typename
```

- b) 在词典 dicta 中添加或删除词

```
for w in words:
    #print(w.word,w.flag)
    dicta=['管理','餐饮','配送','服务','中心','口腔','门诊部','口腔诊所','口腔医院','分公司',
    if (w.flag!="x")&(w.word not in dicta):
        new_word+=w.word
    #print(new_word)
```

- c) 运行程序，得到两个 cut 后的文件

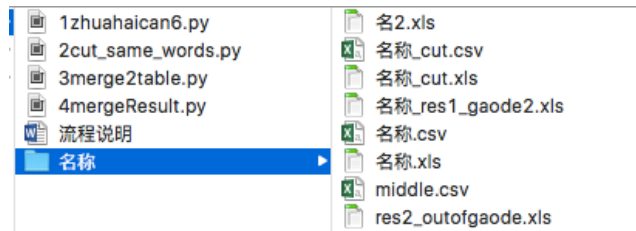


4. 实现两个 cut 表的合并

- a) 在 3merge2table.py 中修改一处名称，并运行程序

```
filename='名称'
dirname=filename+'/'
step3_process_tel(dirname+filename+'.xls',dirname+'middle.csv')
```

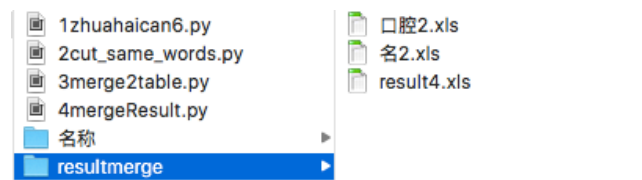
- b) 得到结果



其中，“名 2.xls”为最终给客户文件，“名称\_res1\_gaode2.xls”为我们自己保留的结果文件，“middle.csv”为不重要的中间文件，“res2\_outofgaode.xls”为删除掉高德爬取结果的企查查上的数据

5. 实现结果的合并以及去重

将“名 2.xls”以及的结果文件放到 resultmerge 文件夹中，运行程序即可，结果如图。



其中，result4.xls 为最终合并去重后的结果，4 为总的公司数量

附：

Key

2fd5911c14baa6fe96559c40681f8909

4025623f14283b15cc2c4b452cc3052d

dee4a77852137fdd345d83fa480f7fd1

0d0f94348a01cedc7d1c96e9f871894f

0864776b66cb3ca0f259a2a737bce24d  
82fce00bed5b3ff4791e388a65e6aab0  
94e41f2252e5954b99ee25915b9821e0  
125f11ce12a500fab16774280332f4b7  
f70982ea7aec8e9fe4cd8c3bbd99625  
8f488a2478e519ee74ad4cec185804ab