**市场部收集客户信息和SOP数据比对操作备忘录**

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 内容 | 日期 | 修改者 |
| 1 | 文档初次建立 | 2022-8-4 | Ryan.Wang |

**假设工作目录为fine，所有文件存放在该目录**

**简化步骤：**

1. 按照下面“市场名称收集文件”要求准备好名称文件，例如2018BJAWS.xlsx。
2. 在命令窗中进入工作目录：
3. 在命令窗运行命令：

python fine.py 2018BJAWS.xlsx ditc.txt 现有客户锁定去重.xlsx

1. 将2018BJAWS.xlsx文件拷贝走，再进行适当人工筛选

按需进行其他操作

1. **字典文件**

**Dict.txt**

*1）说明：*

*该文件主要用于过滤掉一些省份、城市、有限公司、科技、分公司之类的词语。*

*日常可以按需进行维护（直接增删改）。*

*在程序中通过对该文件的调用，使得诸如：*

*太平洋电信股份有限公司*

*深圳太平洋电信股份有限公司*

*太平洋电信*

*太平洋电信（深圳）有限公司*

*可以被识别为同一个公司，太平洋电信*

*2）维护*

* + - *手动增加城市名字，如：成都，成都市，这两个名字是不一样的。*
    - *手动修改，方法同上。*
    - *程序自动提取法维护这里不描述。找Ryan。*

1. **市场名单收集文件**
2. 文件名格式

Excel2010以后版本生成的表格文件，如：\*\*\*\*\*.xlsx（这里\*\*\*表示实际文件名，中文英文，不带空格等特殊符号）

例如：

图形用户界面, 应用程序, Teams

描述已自动生成

1. 表格格式

要有两张表单。

第一张表单内容随意；

第二张表单放具体的用户名称，格式参考如下：

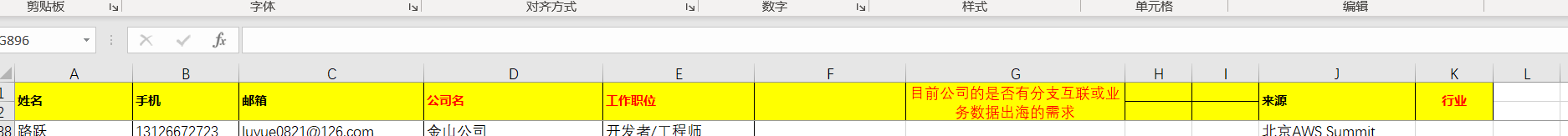
文本, 表格

描述已自动生成

第一张表单内容随意

收集的信息放第二张表单

1. 表单要求



H和I列也为空

F列为空

公司名放在D列

1. 存量客户名称表
2. 文件名格式

Excel2010以后版本生成的表格文件，如：\*\*\*\*\*.xlsx（这里\*\*\*表示实际文件名，中文英文，不带空格等特殊符号）

1. 一张表单即可
2. 客户名称在D列
3. 客户状态在E列
4. 客户简称在F列

说明：

* 截止到2022年8月4日的在网客户已经进行名称简化，以后新增的客户可以直接在F列进行简化。
* 如果存量客户表需要重新生成，建议先去掉特殊符号如数字，连接号-等等。括号可以不去除。并且进行去重处理，以增加精确度。

名称表样表如下图所示：



1. 日常使用

1）在命令窗运行命令： python fine.py 2018BJAWS.xlsx ditc.txt 现有客户锁定去重.xlsx

python： python指令

fine.py: 主程序

2018BJAWS.xlsx： 存放收集客户名称的文件名

dict.txt: 字典文件

现有客户锁定去重.xlsx: 现网客户名单存放文件



备注：需要先安装python运行环境

2）运行结果

* 条目显示：

图片包含 矩形

描述已自动生成

* 命中或者较为匹配时候会有提醒，如下图：

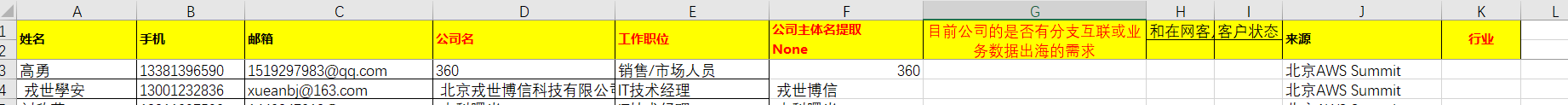
文本

描述已自动生成

* 总命中显示



* 收集客户的名单表格内容显示



和原表格相比，在原来为空的F,H和I列上已经有了内容

建议对H列进行筛选。

图形用户界面, 应用程序, 表格

描述已自动生成图形用户界面, 文本, 应用程序

描述已自动生成

在F列中，显示的是公司的主体名

在H列中，如果是在网客户则会用红色填充，备注为“在网客户”并在I列中显示是“现有客户”还是“锁定客户”。如下图：



* 说明：

1. 准确性依赖于dict.txt字典文件
2. H列中显示“建议人工比对客户”是指该公司主体名和公司存量客户的主体名很相近，建议人工再排查一下。

至此已经完成搜集客户排查

1. 进阶操作
   1. 字典生成

通过以下命令行来生成字典

**Python MakeDict.py test\_w.xlsx dic.txt 30**

*Python: 命令*

*MakeDict.py: 脚本*

*Test\_w.xlsx: 客户名称放在表格文件中的D列，如上面的test\_w.xlsx*

*Dic.txt: 字典文件名，可以不用事先生成，只输入一个名字就可以。*

*30： 表示生成了的字典中有30个词*

生成的字典文件需要人工进行判别，不要误伤（把出现频次比较多的公司主体名给列进去了）

* 1. 存量客户简化名提取和生成

略

1. 可能问题和注意事项

* 运行报错后，会出现表格文件被设成只读模式，这个时候，需要保存其他excel文件，然后在windows中的任务管理器中将excel进程杀掉。
* 表格文件操作之前先保存副本。