代码思路：

> 在这里遇到了很多问题，如：
>
> 1. global company key 和 company name分别提取出作为列表，去重后列表长度却不同了
> 2. 想要将 global company key 和 company name 提取后合成字典写入xls表中，还没有去重字典长度就已经变小了（如下图）
>
> | Name | Type | Size | Value |
> | --- | --- | --- | --- |
> | global_cpkey | list | 65535 | ['001166', '001166', '0… |
> | global_cpnames | list | 65535 | ['ASM INTERNATIONAL NV'… |
> | global_namekey_dict | dict | 29193 | {'ASM INTERNATIONAL NV'… |
> | global_workb | book.Book | 1 | Book object of xlrd.book module |
> | global_worksh | sheet.Sheet | 1 | Sheet object of xlrd.sheet module |
>
> 3. 一开始代码思路不清晰，反复删改了很多次

1. 将global firm 与 us names 中的company name 和 global company key 提取出，并保持原有的对应关系写入新创建的Allcompany .xls表中
2. 将customer表格中Customer Name 与上述中所有的 company name 进行比对后，匹配度大于60% 的则记录在 Allcompany .xls表中,使缩写与全称的公司的对应相同的global company key
3. 将 Allcompany中的company name列 与 customer 表中company name列（供应商列）进行比对，如果名称相同，记录下该供应商对应的customer,再通过在Allcompany中找到对应的customer与Allcompany中的company name 列再进行比对，获取到该customer的全名作为下游，并记录在Allcompany .xls表中
4. 将Allcompany中的company name列 与 customer 表中Customer Name列（客户列）进行比对，如果名称相同，记录下该供应商对应的customer,再通过在customer表中找到该customer对应的供应商作为上游，并记录在Allcompany .xls表中

最后应该呈现的效果：

一行中包括： global company key , company name, customer(下游）,供应商（上游）

> 由于对数据项每一列信息的不熟悉，在总结哪些是可用的信息时花费了一些时间，再加上代码编写过程中思路的不清晰和偶尔出现的bug，今日的代码进展并不顺利。下面是还在删改的第4版，是代码思路和上面相符 （遇到问题还在修改的部分注释了）：

```
import xlrd
import xlwt
import difflib
import functools

# global firm names 和 us names 的 Global Company Key 和 Company Name 提取
global_workb = xlrd.open_workbook(r'C:\Users\jc\Documents\Pydata'+
                                  '\Database Table\global firm names.xls')
global_worksh = global_workb.sheet_by_name('0x77igavdumz8vul')
global_cpnames = global_worksh.col_values(colx = 7, start_rowx = 1)
```

```
global_cpkey = global_worksh.col_values(colx = 0, start_rowx = 1)
# 组成字典
global_namekey_dict ={ name:key for name,key in zip(global_cpnames,
global_cpkey)}

# us_workb = xlrd.open_workbook(r'C:\Users\jc\Documents\Pydata'+
#                                   '\Database Table\\us names.xls')
# us_worksh = us_workb .sheet_by_name('76aqys7wh9axjpme')
# us_cpnames = us_worksh.col_values(colx = 9, start_rowx = 1)
# us_cpkey = us_worksh.col_values(colx = 0, start_rowx = 1)
# # 组成字典
# us_namekey_dict ={name:key for name,key in zip(us_cpnames,us_cpkey)}

# # 对company names去重
# namekey_dict = global_namekey_dict + us_namekey_dict
# unique_list = functools.reduce(lambda x, y: y in x and x or x + [y],
namekey_dict, [])



# # 写入Allcompany表中
# Allcompany = xlwt.Workbook()
# sheet = Allcompany.add_sheet('sheet1')
# name_list = ['Global Company Key','Company Name']
# for i in name_list:
#     sheet.write(0, name_list.index(i), i)



# # 将 global firm names 的 company name和Global Company Key写入xls表中
# Allcompany = xlwt.Workbook()
# sheet = Allcompany.add_sheet('sheet1')
# name_list = ['Global Company Key','Company Name']
# for i in name_list:
#     sheet.write(0, name_list.index(i), i)

# # cpid_dict = dict(zip(global_cpid,range(1,len(global_cpid)+1)))
# # for glocp,i in cpid_dict.items():
# #     sheet.write(i,0,glocp)

# cpname_dict = dict(zip(global_cpnames ,range(1,len(global_cpnames)+1)))
# for glocp,i in cpname_dict.items():
#     sheet.write(i,1,glocp)

# # us names 的 Global Company Key 和 Company Name 提取
# us_workb = xlrd.open_workbook(r'C:\Users\jc\Documents\Pydata'+
#                                   '\Database Table\\us names.xls')
# us_worksh = us_workb .sheet_by_name('76aqys7wh9axjpme')
# us_cpnames = us_worksh.col_values(colx = 9, start_rowx = 1)
# us_cpid = us_worksh.col_values(colx = 0, start_rowx = 1)
# us_cpnames = list(set(us_cpnames))
# us_cpid = list(set(us_cpid))

# # 将us names 的 company name和Global Company Key写入xls表中
# cpid_dict = dict(zip(us_cpid,range(len(global_cpnames)+1,
#                              len(global_cpnames)+1+len(us_cpid))))
# for uscp,i in cpid_dict.items():
```

```
#     sheet.write(i,0,uscp)

# # cpname_dict = dict(zip(us_cpnames ,range(len(global_cpnames)+1,
# #                          len(global_cpnames)+1+len(us_cpnames))))
# # for uscp,i in cpname_dict.items():
# #     sheet.write(i,1,uscp)


# # 将customer中的Customer Name与company name用difflib进行差异对比


# # 保存文件
# Allcompany.save('Allcompany.xls')
```