In [ ]:
```
import xlrd
import pandas as pd
import xlsxwriter
import glob
import os
import re
import numpy as np
import sys
import datetime
```

In [ ]:
```
path1 = 'C:/Users/org81/Dropbox/Data_analysis/seekingalpha/full_indexed_revised.csv'
df1 = pd.read_csv(path1, encoding='utf-8')
df1.count()
```

In [ ]:
```
path1 = 'C:/Users/Rocku/Dropbox/Data_analysis/seekingalpha/index_file.csv'
df1 = pd.read_csv(path1)
df1.count()
```

In [ ]:
```
df_ex = df1.loc[df1['Role'] == 'Executives']
df_ex[:100]
```

In [ ]:
```
file_names = list(set(df_ex['FileName'].tolist()))
```

In [ ]:
```
len(file_names)
```

In [ ]:
```
columns = ['FileName', 'CompanyName', 'Exchange', 'Ticker',
           'CallDate', 'UploadDate', 'Role', 'OriginalPosition', 'PositionDetail', 'N
ame',
           'FileIndex','NameIndex', 'PositionIndex1', 'PositionIndex2']

df = pd.DataFrame(columns=columns)

print('--------------start indexing--------------')

file_index = 90000

for f_index, file_name in enumerate(file_names[90000:]):
    file_index += 1

    if (file_index % 1000 == 1):
        print(file_index, ': Proceeding', datetime.datetime.now())

    df_indi = df_ex[df_ex['FileName'] == file_name]

    name_index = 0
    for person_index, person_row in df_indi.iterrows():
        positions = str(person_row['Position'])
        positions = positions.replace(' and ', ', ')
        positions_lists = re.split(r', ', positions)
        name_index += 1

        for position_index, position_row in enumerate(positions_lists):
            data_dict = {}
            data_dict['FileName'] = person_row['FileName']
            data_dict['UploadDate'] = person_row['UploadDate']
            data_dict['CompanyName'] = person_row['CompanyName']
            data_dict['Exchange'] = person_row['Exchange']
            data_dict['Ticker'] =person_row['Ticker']
            data_dict['CallDate'] = person_row['CallDate']
            data_dict['Role'] = person_row['Role']
            data_dict['Name'] = person_row['Name']
            data_dict['PositionDetail'] = position_row
            data_dict['OriginalPosition'] = positions
            data_dict['FileIndex'] = file_index + 1
            data_dict['NameIndex'] = name_index
            data_dict['PositionIndex1'] = name_index
            data_dict['PositionIndex2'] = position_index + 1


            df_new = pd.DataFrame([data_dict], columns=columns)
            df = df.append(df_new)

print('--------------end indexing--------------')
```

In [ ]:
```
df.to_csv('full_indexed.csv', header='column_names', index=False, encoding='utf-8')
```