

1.2 文件管理

电脑中的文件是日常工作、学习和生活中最常用的数据载体，如何快速地处理文件是财务工作的一项重要技能。Python中的pathlib库，可以帮助我们方便而又快速地定位和操作文件。

In [1]:

```
#import required modules
from pathlib import Path
import pandas as pd

#display option in jupyter notebook
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
pd.set_option('max_colwidth', 100)
```

在Python中表示文件或文件夹的路径

Windows系统中，常见的文件路径如：C:\Users\Zhangsan\Documents\report.docx；文件夹路径则如：D:\Courses\Accounting。

在Python中，可以使用字符串或Path对象来表示这些路径。

In []:

```
# Use a string to denote a path
file = 'D:/Data/report_data/2014/Alibaba_bs.csv'
folder = 'D:/Data/report_data/2014'
```

In [13]:

```
# Use a Path object to represent a path
file_obj = Path('D:/Data/report_data/2014/Alibaba_bs.csv')
folder_obj = Path('D:/Data/report_data/2014')
```

In [19]:

```
# 路径对象的属性
file_obj.parent
```

Out[19]:

```
WindowsPath('D:/Data/report_data/2006')
```

打开一个文件

In []:

```
# 向open函数传入一个字符串
with open(file) as f:
    temp_text = f.read()
print(temp_text)
```

In [15]:

```
# 向open函数传入一个Path对象
with open(file_obj) as f:
    temp_text = f.read()
print(temp_text)
```

some content

遍历文件夹

所谓遍历，即访问到文件夹中的每个元素（文件夹或文件）

In [7]:

```
files = list(folder_obj.iterdir())
files
```

Out[7]:

```
[WindowsPath('D:/Data/report_data/2006/Alibaba_bs.csv'),
 WindowsPath('D:/Data/report_data/2006/Alibaba_cf.csv'),
 WindowsPath('D:/Data/report_data/2006/Alibaba_is.csv'),
 WindowsPath('D:/Data/report_data/2006/Amazon_bs.csv'),
 WindowsPath('D:/Data/report_data/2006/Amazon_cf.csv'),
 WindowsPath('D:/Data/report_data/2006/Amazon_is.csv'),
 WindowsPath('D:/Data/report_data/2006/Apple_bs.csv'),
 WindowsPath('D:/Data/report_data/2006/Apple_cf.csv'),
 WindowsPath('D:/Data/report_data/2006/Apple_is.csv'),
 WindowsPath('D:/Data/report_data/2006/Gucci_bs.csv'),
 WindowsPath('D:/Data/report_data/2006/Gucci_cf.csv'),
 WindowsPath('D:/Data/report_data/2006/Gucci_is.csv'),
 WindowsPath('D:/Data/report_data/2006/Tencent_bs.csv'),
 WindowsPath('D:/Data/report_data/2006/Tencent_cf.csv'),
 WindowsPath('D:/Data/report_data/2006/Tencent_is.csv'),
 WindowsPath('D:/Data/report_data/2006/test.txt')]
```

【练习2-1】统计某个文件夹下（D:/Data/report_data/2024/）的文件有哪几种扩展名，以及每种扩展名有几个文件。

In [9]:

```
f_types = {}

for f in folder_obj.iterdir():
    if f.suffix not in f_types:
        f_types[f.suffix] = 1
    else:
        f_types[f.suffix] += 1

f_types
```

Out[9]:

```
{'.csv': 15, '.txt': 1}
```

In [5]:

```
#找出满足特定条件的文件
#通配符: ?代表一个字符, *代表任意个字符。
files = list(folder_obj.glob('Ali*.csv'))
files
```

Out[5]:

```
[WindowsPath('D:/Data/report_data/2006/Alibaba_bs.csv'),
 WindowsPath('D:/Data/report_data/2006/Alibaba_cf.csv'),
 WindowsPath('D:/Data/report_data/2006/Alibaba_is.csv')]
```

查看当前文件夹

In []:

```
current_directory = Path.cwd()
current_directory
```

访问文件夹及其子文件夹中的所有文件

【练习2-2】将某文件夹（D:/Data/reportdata）里的子文件夹下的所有csv文件重新命名，在原有文件名中间加入“[上级文件夹]”的文字。如，原先在 D:/Data/report_data/2024文件夹下的“Alibaba_cs.csv”文件，重命名为“Alibaba_2024_cs.csv”。

In [22]:

```
folder_obj = Path('D:/Data/report_data')

for p in folder_obj.rglob('*.csv'):
    p.rename(p.parent/(p.name[:-7]+'_'+p.parts[-2]+p.name[-7:]))
```