

结构预览

```
1  -- 4-vgtime
2      -- data # 存放全部游戏封面和游戏详情信息
3          -- Images
4          -- final_details.csv
5          -- first_game_info.csv
6
7      -- docs # 需求文档
8      -- input_files # 存放需要的初始excel和csv文件
9          -- A1_主播表.xlsx
10         -- NewGameList.csv #所有游戏中文名称 用来进行搜索爬虫
11         -- final_details # 为上述data文件夹中的拷贝文件
12         -- 新用户名表.xlsx
13     -- log # log文件夹, 按照运行程序的日期建立相应log文件夹
14     -- output_files # 输出文档
15         -- A1_主播表.csv
16         -- new_details.csv # 将原来的final_details.csv文件中的日期进行更改调整
17     -- src # 存放所有代码文件
18         -- step1_vgtime_parse_2.py # 爬取数据
19         -- step2_change_anchor_name.py # 更改主播表中的用户名
20         -- step3_search_stream.py # 将平台数据写入到主播表
21     -- gamelist20210830.xlsx # 将此文件保存成csv放入input_files文件夹中。
```

运行步骤

爬取游戏信息

1. 将gamelist20210830.xlsx保存成 NewGameList.csv 文件放入input_files文件夹。
2. 将 step1_vgtime_parse_2.py 中 main_process 函数内的 offsets 设置成相应的数字:

- ```
1 - 比如我们有89个游戏要爬取, 分为四份, 可以设置为0, 25, 50, 75
2 - 那么`offsets = [0, 25, 50, 75]`
```

3. 运行 step1\_vgtime\_parse\_2.py 文件即可。

```
def main_process():
 """Since we use multiprocessing for this programme, the order of the game is not so sequential, however, we
 use the original chinese name as the index to instruct us."""
 log_folder_path = create_log_folder()

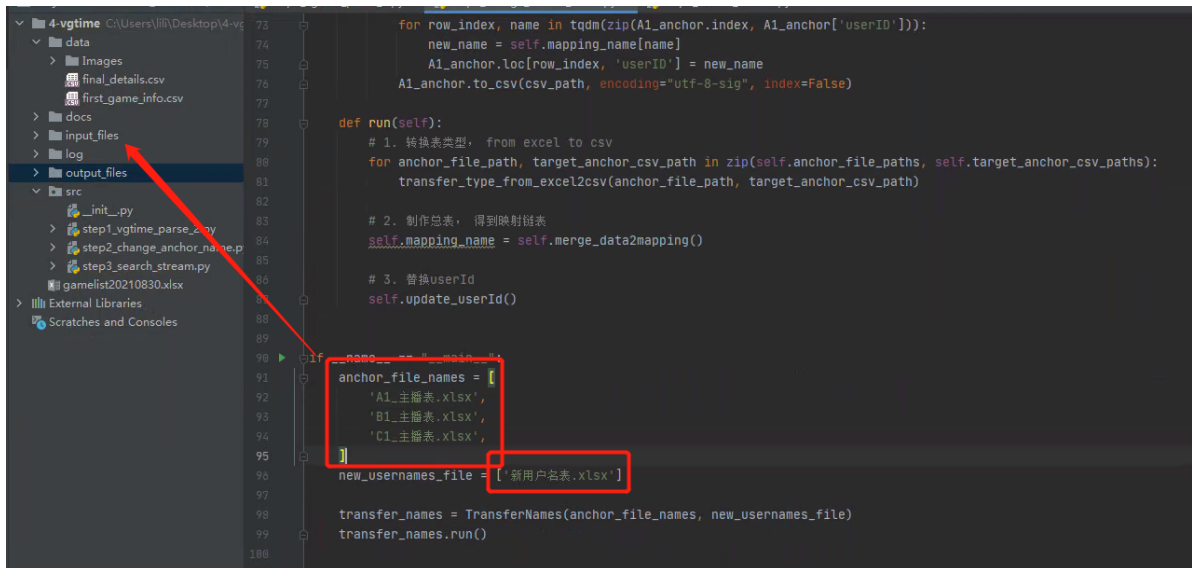
 game_list_data = pd.read_csv(os.path.abspath(os.path.join(os.getcwd(), '..', 'input_files', 'NewGameList.csv')))

 # game_name_data = game_list_data.loc[:, 'GameName'][offsets[0]:offsets[1]]
 offsets = [0, 1000, 2000, 3000]
 game_name_lists = [game_list_data.loc[:, 'GameName'][offsets[0]:offsets[1]],
 game_list_data.loc[:, 'GameName'][offsets[1]:offsets[2]],
 game_list_data.loc[:, 'GameName'][offsets[2]:offsets[3]],
 game_list_data.loc[:, 'GameName'][offsets[3]:]]

 # print(game_name_list)
 with concurrent.futures.ProcessPoolExecutor(max_workers=4) as executor:
 for offset, game_name_list in zip(offsets, game_name_lists):
 executor.submit(main_thread, offset, game_name_list, log_folder_path)
```

## 替换主播表中userId

1. 将主播表与新用户名表存放在 input\_files 文件夹中。
2. 在 step2\_change\_anchor\_name.py 文件中，下图位置的 anchor\_file\_names 将上述文件夹中需要的所有主播表的文件名写入。
3. 如果新用户名表名称未变，在下图代码红色框中则不需要改变。
4. 运行该程序即可。



## 将平台写入主播表

1. 更改需要写入的主播表名称写入下方途中的红色方框。

