



链滴

表格数据提取

作者: [cuijianzhe](#)

原文链接: <https://ld246.com/article/1608017828070>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

```

#!/usr/bin/python
# -*- coding: utf-8 -*-
# @Time : 2020/12/14 13:07
# @Author : cuijianzhe
# @File : xlwt.py
# @Software: PyCharm

import openpyxl
import requests
import os
import random
name = input('请输入表格名称:')
path = name + '.xlsx'
wb = openpyxl.load_workbook(path)

sheetnames = wb.sheetnames #获取文件中所有的sheet
sheetData = wb[sheetnames[0]] #第一个sheet的内容

rows_max = sheetData.max_row #最大行数
cols_max = sheetData.max_column #最大列数

ApprovalNumber = sheetData.cell(row = 1,column = 4).value
def get_case(): #获取所有表头功能
    case_list = []
    for cols in range(1,cols_max): #列数
        cases = sheetData.cell(row=1,column=cols).value
        case_list.append(cases.replace('\n\n','\n'))
    return case_list

def generate_code(char):
    all_char = '0123456789qazwsxedcrfvtgbyhnujmikolpQAZWSXEDCRFVTGBYHNUJIKOLP'
    index = len(all_char) - 1
    code = ""
    for _ in range(int(char)):
        num = random.randint(0,index)
        code += all_char[num]
    res = "".join(code)
    return res

def get_carNum():
    ''' 提取并添加站点名称 '''
    # 第一个方法使用列表
    cars_list = [] #存储车牌号
    try:
        for cols in range(1,cols_max):
            carNum = sheetData.cell(row = 1,column = cols).value #每列的值
            if carNum == '车牌号':
                for rows in range(2,rows_max+1): #遍历最大行数
                    carnumName = sheetData.cell(row = rows,column=cols).value
                    if carnumName != "":
                        cars_list.append(carnumName)
                    else:
                        cars_list.append("空车牌号在第%s行"%(rows))
    except Exception as error:

```

```

    print('get车牌号错误', 'error:{}'.format(error))
    return cars_list

def get_url(case_list):
    url_list = []

    for cols in range(1, cols_max): # 列数
        datas = sheetData.cell(row=1, column=cols).value
        if datas == case_list:
            for rows in range(2, rows_max+1):
                images_url = sheetData.cell(row=rows, column=cols).value
                if images_url != "":
                    url_list.append(images_url.replace("\n\n", '\n'))
                else: #如果此单元格可能为空的话, 添加自定义链接占位
                    url_list.append('https://file.cjzshilong.cn/pictures_file/guohui-e67e7b3b.png\n')
            else:
                continue
    return url_list

def Download(url, car_name):
    for num in range(len(car_name)):
        ApprovalNumber = sheetData.cell(row=num+2, column=1).value # 订单编号
        name = car_name[num] #车牌目录
        name_dir = car_name[num] + '_' + ApprovalNumber #同车牌 不同订单 创建目录
        list_files = os.listdir('./')
        if name not in list_files:
            os.mkdir(name)
            os.chdir(name)
            os.mkdir(name_dir)
            os.chdir(name_dir)
            file_list = url[num].split('\n')
            with open('./统计' + name_dir + '_' + generate_code(4) + '的图片url_共计' + str(len(file_
ist)) + '条.txt', 'w', encoding='utf-8') as f:
                f.write(url[num])

            for url_car in file_list:
                if url_car != "":
                    response = requests.get(url_car)
                    with open(name + '_' + generate_code(15) + '.jpg', 'wb') as car:
                        car.write(response.content)
                    print('%s下载完毕' %name_dir)

            os.chdir('../..')
        else:
            namedir_list = os.listdir('./' + name)
            os.chdir(name)
            if name_dir not in namedir_list:
                os.mkdir(name_dir)
                os.chdir(name_dir)
                file_list = url[num].split('\n')
                with open('./统计' + name_dir + '_' + generate_code(4) + '的图片url_共计' + str(len(fi
e_list)) + '条.txt', 'w', encoding='utf-8') as f:
                    f.write(url[num])

```

```

for url_car in file_list:
    if url_car != "":
        response = requests.get(url_car)
        with open(name + '_' + generate_code(15) + '.jpg', 'wb') as car:
            car.write(response.content)
            print('%s下载完毕' % name_dir)
    os.chdir('../..')
else:
    os.chdir(name_dir)
    file_list = url[num].split('\n')
    with open('./统计' + name_dir + '_' + generate_code(4) + '的图片url_共计' + str(len(file_list)) + '条.txt', 'w', encoding='utf-8') as f:
        f.write(url[num])

for url_car in file_list:
    if url_car != "":
        response = requests.get(url_car)
        with open(name + '_' + generate_code(15) + '.jpg', 'wb') as car:
            car.write(response.content)
            print('%s下载完毕' % name_dir)
    os.chdir('../..')

if __name__ == '__main__':
    case_list = ['车辆正面照&车辆安全检查照','司机安全照','起始用车仪表盘公里数','结束用车仪表盘公里数','路桥费照片','停车费照片']
    car_name = get_carNum()
    for func in case_list:
        url_img = get_url(func)
        Download(url_img, car_name)

```