

通过 python 读取 excel 表格内容 url 批量 下载

作者: [cuijianzhe](#)

原文链接: <https://ld246.com/article/1578120697736>

来源网站: [链滴](#)

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


```

Data_sheet = workbook.sheets()[0]
rowNum = Data_sheet.nrows #行数
colNum = Data_sheet.ncols #列数

def get_sitesname():
    """ 提取并添加站点名称 """
    #第一个方法使用列表
    sites_list = []
    i = 1 #从1开始跳过表头数据
    try:
        for s in range(Data_sheet.ncols):
            if (Data_sheet.cell_value(0, s)) == '站点名称':
                while i <= rowNum:
                    if Data_sheet.cell_value(i,s) != "":
                        sites_list.append(Data_sheet.cell_value(i,s)+"_"+str(i))
                    i += 1
    except:
        pass
    return sites_list
    #第二个方法使用列表推导式直接返回
    # return [Data_sheet.cell_value(i, 3) + "_" + str(i) for i in range(1,2) if i not in [] for i in range(
, rowNum-2) if Data_sheet.cell_value(i, 3) != ""]

def get_url():
    """ 提取链接所有内容 提取对应行列的内容 """
    url_list = []
    try:
        for b in range(Data_sheet.ncols):
            if (Data_sheet.cell_value(0,b)) == '图片地址':
                for row in range(1,rowNum-2):
                    if Data_sheet.cell_value(row,b) != "":
                        url_list.append(Data_sheet.cell_value(row, b).replace(';', '\n'))
                    else:
                        url_list.append('https://file.cjzshilong.cn/pictures_file/guohui-e67e7b3b.png\n')
    except:
        pass

    return url_list

def Download(url,site_name):
    """ 构建目录 批量下载图片家目录有问题。待解决优化 """
    for num in range(len(url)):
        name = site_name[num]
        os.mkdir(name)
        os.chdir(name)
        file_list = url[num].split('\n')
        with open('./' + name[:-2] + '图片url_共计'+ str(len(file_list)-1) + '条.txt', 'w', encoding='ut
-8') as f:
            f.write(url[num])
            d = 0
            for url_1 in file_list:
                if url_1 != "":
                    res = requests.get(url_1)
                    with open(name[:-2] + '_' + str(d+1) + '.jpg', 'wb') as p:
                        p.write(res.content)
                    print('%s下载完毕' %name)

```

```

        d += 1
    os.chdir('..')

if __name__ == '__main__':
    url_ss = get_url()
    site_name = get_sitesname()
    Download(url_ss,site_name)

```

第三版：

- 根据图片找到对应站点名称
- 每个站点生成单独的目录

```

#!/usr/bin/python
# -*- coding: utf-8 -*-
# @Time : 2019/10/15 21:36
# @Author : cuijianzhe
# @File : biaoqe.py
# @Software: PyCharm
import xlrd
import requests
import os
path = 'test.xls'
workbook = xlrd.open_workbook(path)
Data_sheet = workbook.sheets()[0] #代表表格中sheet 0表格内容
rowNum = Data_sheet.nrows #行数
colNum = Data_sheet.ncols #列数

'''
提取单元格所有内容
'''

# print(Data_sheet.cell_value(1,4)) #第1行第4列单元格内容
# print(Data_sheet.ncols) #有效列数

sites_list = []
def get_sitesname():
    '''
    提取并添加站点名称
    '''
    i = 0
    try:
        while i <= rowNum:
            sites_list.append(Data_sheet.cell_value(i,3))
            i += 1
    except:
        pass
    del sites_list[0]
    del sites_list[-1]
    del sites_list[-1]
    '''
    所有站点名称加上后缀,达到不重复效果
    '''
    s = 0

```

```

for i in sites_list:
    s += 1
    weizhi = sites_list.index(i)
    sites_list[weizhi] = i + '_' + str(s)
return sites_list
def get_url():
    """
    提取链接所有内容
    提取对应行列的内容
    """
    name_col = '图片地址'
    url_list = []
    try:
        for b in range(Data_sheet.ncols):
            if (Data_sheet.cell_value(0,b)) == name_col:
                for col in range(colNum):
                    url_data = Data_sheet.cell_value(col,b)
                    url_data2 = url_data.replace(';', '\n')
                    url_list.append(url_data2)
    except:
        pass
    del url_list[0]
    del url_list[-1]
    del url_list[-1]
    """
    把列表中元素为空的值替换成自己定义的
    """
    for i in range(int(len(url_list))):
        if url_list[i] == "":
            url_list[i] = 'https://file.cjzshilong.cn/pictures_file/guohui-e67e7b3b.png'
    return url_list

def Download(url,site_name):
    """
    构建目录
    批量下载图片家目录有问题。待解决优化
    """
    c = 0
    for num in range(len(url)):
        name = site_name[num]
        os.mkdir(name)
        os.chdir(name)
        with open('./url_' + str(c) + '.txt', 'w', encoding='utf-8') as f:
            f.write(url[num])
        with open('./url_' + str(c) + '.txt', 'r', encoding='utf-8') as d:
            file = d.read()
            file_list = file.split("\n")
            name = site_name[num]
            c += 1
            d = 0
            for url_1 in file_list:
                if url_1 == "":
                    break
                res = requests.get(url_1)

```

```

        print(('下载完毕') % (name))
        with open(name + str(d) + '.jpg', 'wb') as p:
            p.write(res.content)
        d += 1
    os.chdir('..')

if __name__ == '__main__':
    url_List = get_url()
    site_name = get_sitesname()
    Download(url_List,site_name)

```

1.2 异常捕获

表格中一个单元格中有一个 url 或者多个 url 或者空白，解决不识别情况，str 转 list 解决

```

#!/usr/bin/python
# -*- coding: utf-8 -*-
# @Time : 2019/10/15 21:36
# @Author : cuijianzhe
# @File : biaoage.py
# @Software: PyCharm
import xlrd
import requests
import random
import string
import os
import time
path = 'test.xls'
date = time.strftime('%Y-%m-%d',time.localtime())
workbook = xlrd.open_workbook(path)
# print(workbook.sheet_names())
Data_sheet = workbook.sheets()[0]
# print(Data_sheet.name)
rowNum = Data_sheet.nrows #行数
colNum = Data_sheet.ncols #列数
'''
提取单元格所有内容
'''
list = []
for i in range(rowNum):
    rowlist = []
    for j in range(colNum):
        rowlist.append(Data_sheet.cell_value(i,j))
    list.append(rowlist)

'''
提取对应行列的内容
'''
#print(Data_sheet.cell_value(0,43)) #列名称
#print(Data_sheet.ncols) #有效列数
name_col = '图片地址'
url_list = []
try:
    for b in range(Data_sheet.ncols):

```



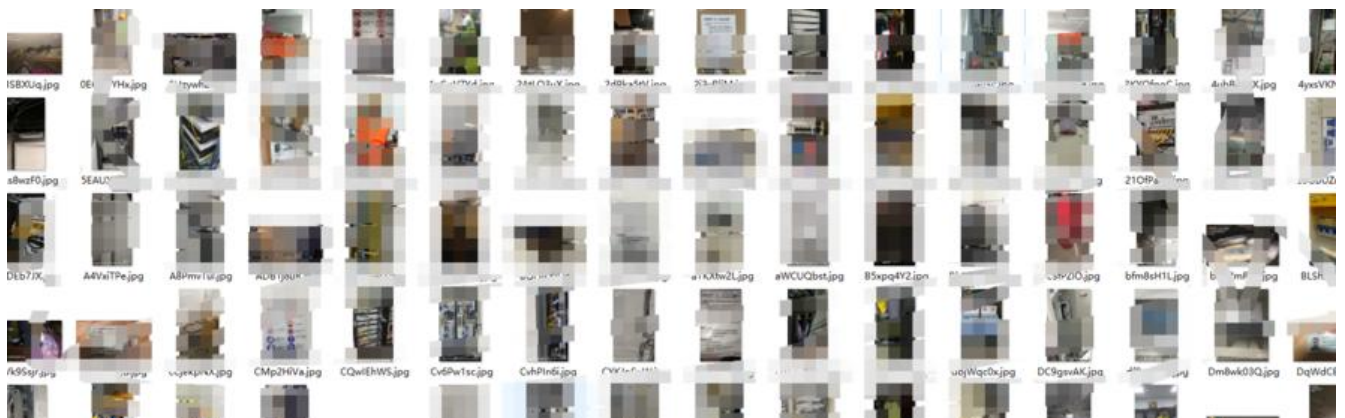
```

        if (Data_sheet.cell_value(0,b)) == name_col:
            for col in range(colNum):
                url_data = Data_sheet.cell_value(col,b)
                url_data2 = url_data.replace(';','\n')
                url_list.append(url_data2)
except:
    pass

del url_list[0]
with open('./url.txt','w',encoding='utf-8') as f:
    for url in url_list:
        f.write(str(url))
        # f.write(',')
with open('./url.txt','r',encoding='utf-8') as d:
    file = d.read()
    file_list = file.split('\n')
del file_list[-1]

headers = {
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like
Gecko) Chrome/75.0.3770.142 Safari/537.36'
}
name = 'pictures'
os.mkdir(name)
os.chdir(name)
for url_1 in file_list:
    name_1 = ''.join(random.sample(string.ascii_letters + string.digits, 8))
    #name_1 = url_1.replace('/', '')
    res = requests.get(url_1,headers=headers)
    print((' %s 下载完毕' )%(name_1))
    with open(name_1 + '.jpg','wb') as p:
        p.write(res.content)

```

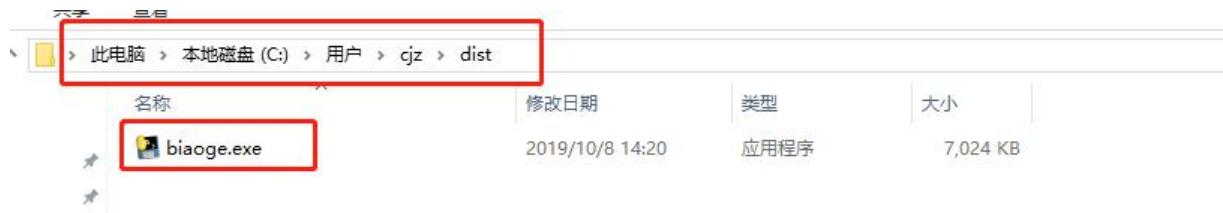


- 安装 `pyinstaller`

- pyinstaller 的使用

pyinstaller -F biaoge.py

- 找到文件



参考:

[pyinstaller官方文档:](#)

[xlrd官方文档:](#)

<https://www.cnblogs.com/insane-Mr-Li/p/9092619.html>