

个性网上爬取照片

笔记本: reptile_draft

创建时间: 2021/6/12 13:37

更新时间: 2021/6/12 13:37

作者: 134exetj717

URL: about:blank

```
# 请求网页
import time
import urllib
from bs4 import BeautifulSoup
import requests
import re

#找到图片资源
findPic = re.compile('<a href=".*?" target="_blank">\n\n<span class="tit">(.*?)</span>\n</a>')
#取出图片
def getWeb():
    header = {
        'Accept':
        'text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8',
        'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:89.0)
Gecko/20100101 Firefox/89.0'
    }
    #获取网页
    #response = requests.get('https://www.vmgirls.com/15198.html',
headers=header)
    url = 'https://www.bizhizu.cn/nvsheng/weimei/'
    data = bytes(urllib.parse.urlencode({'name': 'eric'}), encoding='utf-8')
    dx = urllib.request.Request(url=url, headers=header, data=data)
    response = urllib.request.urlopen(dx)
    html = response.read().decode('utf-8')
    return html

#解析网页
def analyseWeb(baseHtml):
    content = BeautifulSoup(baseHtml, 'html.parser')
    urls = []
    for item in content.find_all('a'):
        pic = re.findall(findPic, str(item))
        if len(pic) != 0:
            urls.append(pic[0])
    return urls

#保存图片
def savePic(urls):
    header = {
        'Accept':
        'text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8',
        'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:89.0)
Gecko/20100101 Firefox/89.0'
    }
    for url in urls:
        file_name = './pic/' + url[1] + '.jpg'
```

```
        response = requests.get(url[0], headers=header)
        f = open(file_name, 'wb')
        f.write(response.content)
        f.close()

if __name__ == '__main__':
    baseHtml = getWeb()
    urls = analyseWeb(baseHtml)
    print(urls)
    savePic(urls)
```