etree练习之4k图片在线下载

笔记本:	reptile_draft		
创建时间:	2021/7/11 15:57	**更新时间:**	2021/7/11 15:57
作者:	134exetj717		
URL:	about:blank		

```
# -*- codeing = utf-8 -*-
# @Time : 2021/7/11 14:37
# @Author : 希杰
# @file : 4k图片在线下载.py
# @software : PyCharm


#需求：解析下载图片数据 https://pic.netbian.com/4kmeinv


import requests
from lxml import etree
import os
if __name__ == '__main__':
    url = 'https://pic.netbian.com/4kmeinv/'
    headers = {
        'User-Agent': 'Mozilla / 5.0(Windows NT 10.0;  Win64;  x64; rv: 89.0)
Gecko / 20100101  Firefox / 89.0'
    }
    # response = requests.get(url=url,headers=headers)
    # # #手动设定响应数据的编码格式，可能解决不了乱码问题
    # # response.encoding = 'utf-8'
    # # page_text = response.text
    #如果用text容易出现中文乱码，content往往能正确显示
    page_text = requests.get(url=url,headers=headers).content
    #数据解析：src的数据值、alt属性
    tree = etree.HTML(page_text)
    li_list = tree.xpath('//div[@class="slist"]/ul/li')
    path_name = './bian'
    if not os.path.exists(path_name):
        os.mkdir(path_name)
    for li in li_list:
        img_src = 'https://pic.netbian.com/' + li.xpath('./a/img/@src')[0]
        #当中文出现乱码时，说明编码格式不一样
        img_name = li.xpath('./a/img/@alt')[0] + '.jpeg'
        print(img_src,img_name)
        # #通用处理中文乱码的解决方案
        # img_name.encode('iso-8859-1',)
        img_text = requests.get(url=img_src,headers=headers).content
        print(img_text)
        if os.path.exists(path_name + '/'+img_name):
            os.mkdir(path_name + '/'+img_name)
        fp = open(path_name + '/'+img_name,'wb')
        fp.write(img_text)
    print('爬取成功！')
```