



链滴

python 爬虫入门，单页面爬取图片

作者: [WhiteLoong](#)

原文链接: <https://ld246.com/article/1535599470028>

来源网站: 链滴

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

```
<h2 id="导入相应模块">导入相应模块</h2>
<p>import requests<br>
import urllib.request<br>
from lxml import etree</p>
<h2 id="请求头部-模拟浏览器访问">请求头部，模拟浏览器访问</h2>
<p>headers = {<br>
'User-Agent': 'Mozilla/5.0 (Linux; Android 6.0; Nexus 5 Build/MRA58N) AppleWebKit/537.36 (
HTML, like Gecko) Chrome/68.0.3423.2 Mobile Safari/537.36'<br>
}</p>
<h2 id="爬取图片的访问地址">爬取图片的访问地址</h2>
<p>url = '<a href="https://ld246.com/forward?goto=http%3A%2F%2Fwww.meizitu.com%2F
%2F5592.html" target="_blank" rel="nofollow ugc">http://www.meizitu.com/a/5592.html</a
'</p>
<h2 id="使用requests请求-访问-网页">使用 requests 请求（访问）网页</h2>
<p>html = requests.get(url,headers=headers).content</p>
<h2 id="使用etree解析整个网页">使用 etree 解析整个网页</h2>
<p>xp_html = etree.HTML(html)</p>
<h2 id="使用xpath定位到需要爬取的图片名称以及连接">使用 xpath 定位到需要爬取的图片名称
及连接</h2>
<p>imgnames = xp_html.xpath('//div/p/img/@alt')<br>
imgurls = xp_html.xpath('//div/p/img/@src')</p>
<h2 id="循环下载连接中的图片并按照名称保存到D盘XX文件夹下">循环下载连接中的图片并按照
称保存到 D 盘 XX 文件夹下</h2>
<p>for (imgname,imgurl) in zip(imgnames,imgurls):<br>
try:<br>
urllib.request.urlretrieve(imgurl,'D:\XX\%s.jpg' % imgname)<br>
except Exception as e:<br>
print(imgname + ': 下载出错，地址为: '+ imgurl)</p>
<p>print('下载完成')</p>
```