

## 湖 南 石 化 职 院 教 案

日期	2020 年 4 月 21 日	周次	第八周	时数	4 课时
教学单元	批量获取网络图片数据				
教学目标	<p>知识目标：</p> <ol style="list-style-type: none"> <li>1. 掌握 Python 程序的设计和编程能力；</li> <li>2. 了解大批量获取网络图片的原理；</li> <li>3. 掌握使用 Python 语言批量获取网络图片。</li> </ol>				
	<p>技能目标：</p> <ol style="list-style-type: none"> <li>1. 能够进行大批量获取网络图片的原理，并掌握批量获取网络图片的方法；</li> <li>2. 能够用网络爬虫中的 urllib、urllib2 等模块对图虫网站中的图片进行批量下载，并存储在指定的文件夹中。</li> </ol>				
	<p>素质目标：</p> <ol style="list-style-type: none"> <li>1. 树立正确的政治素质（爱党爱国、品行端正、遵纪守法）；</li> <li>2. 培养团队协作意识；培养创新能力；</li> <li>3. 树立正确的职业目标。</li> </ol>				
教学重点	<ol style="list-style-type: none"> <li>1. 掌握 Python 程序的设计和编程能力；</li> <li>2. 掌握使用 Python 语言批量获取网络图片。</li> </ol>				

教学难点	1. 掌握 Python 程序的设计和编程能力； 2. 掌握使用 Python 语言批量获取网络图片。
教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>1. 新课引入</p> <p>2. 知识讲解</p> <p>1. 设计批量获取图片的程序</p> <p>分析实验要求</p> <p>批量获取并存储网络图片</p> <p>记录程序代码</p> <p>记录并分析实验结果。</p> <p>批量获取网络图片的方法：</p> <p>2. 批量获取网络图片的方法是通过解析网页的 HTML 文件，利用正则表达式把源代码中的图片地址过滤出来，从而根据过滤出来的图片地址下载网络图片。具体来说，批量获取网络图片的方法分为三种，一是用微软提供的扩展库 win32com 来操作 IE，二是用 selenium 的</p>	

webdriver, 三是用 python 自带的 HTMLParser 解析。

实现代码如下:

```
#获取二级页面 url
```

```
def findUrl2(html):
```

```
    re1=r' http://tuchong.com/\d+/\d+|http://\w+(?<!photos).
```

```
tuchong.c    om/\d+/'
```

```
    url2list = re.findall(re1,html)
```

```
    url2lstfltr = list(set(url2list))
```

```
    url2lstfltr.sort(key=url2list.index)
```

```
    #print url2lstfltr
```

```
    return url2lstfltr
```

```
#获取 html 文本
```

```
def getHtml(url):
```

```
    html = urllib2.urlopen(url).read().decode('utf-8')#解码为  
utf-8
```

```
    return html
```

```
#下载图片到本地
```

```
def download(html_page, pageNo):
```

```
    #定义文件夹的名字
```

```
    x = time.localtime(time.time())
```

```
    foldername=str(x.__getattr__("tm_year"))+"-"+str(x.
```

```
__getattr
```

```

        ibute__("tm_mon"))+"-"+str(x.__getattribute__("tm_mday")
)

re2=r' http://photos.tuchong.com/.+/f/.+\.jpg'

imglist=re.findall(re2,html_page)

print imglist

download_img=Nonefor imgurl in imglist:

picpath='C:\\Users\\peterlindi\\Desktop\\lindi\\%s\\%s' %
(foldername, str(pageNo))

filename = str(uuid.uuid1())

if not os.path.exists(picpath):

    os.makedirs(picpath)

    target= picpath+"\\%s.jpg" % filename

    print "The photoslocation is:"+target

    download_img = urllib.urlretrieve(imgurl, target)

    time.sleep(1)

    print(imgurl)

return download_img

```

### 3. 课堂总结

Python 程序的设计和编程能力；

大批量获取网络图片的原理；

使用 Python 语言批量获取网络图片。

参考资料 及应用	《Python3 网络爬虫开发实战》 主编：崔庆才 出版社：人民邮电出版社
作业	完成批量获取网络图片数据
教学反思	