

	2020 4 21				4
	<p style="text-align: center;">:</p> <ol style="list-style-type: none"> 1. Python ; 2. 了 ; 3. Python 。 				
	<p style="text-align: center;">:</p> <ol style="list-style-type: none"> 1. , 2. 中 urllib、urllib2 中 下 , 中。 				
	<p style="text-align: center;">:</p> <ol style="list-style-type: none"> 1. (、 、) ; 2. ; ; 3. 业 。 				
	<ol style="list-style-type: none"> 1. Python ; 2. Python 。 				

1. Python ;

2. Python 。

;

云 ; ;

1.

2.

1.

```

webdriver, 三      python      HTMLParser      。

    代    下:

#    二      url

def findUrl2(html):

    rel=r' http://tuchong.com/\d+/\d+/?http://\w+(?!photos).
tuchong.c    om/\d+/'

    url2list = re.findall(rel,html)

    url2lstfltr = list(set(url2list))

    url2lstfltr.sort(key=url2list.index)

    #print url2lstfltr

    return url2lstfltr

#    html

def getHtml(url):

    html = urllib2.urlopen(url).read().decode('utf-8')#    为
utf-8

    return html

#下

def download(html_page, pageNo):

    # 义
    x = time.localtime(time.time())

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

```

```
ibute__m_m())+str(pageNo)+str(pageNo)
)
re2=r'http://photos.tuchong.com/.*?/.+\.jpg'
imglist=re.findall(re2,html_page)
print imglist
download_img=None
for 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
```

	《Python3 主 : : 人
业	