

-----用户指南-----

参数名 (args)	说明 (instructions)
folder_name *	该参数表示：自定义存储 html 文件的根目录名称，即文件夹名称。
url *	该参数表示：url 前半部分，即，页码数字之前的部分。如：https://coinmarketcap.com/2/ 的 url 前半部分为数字 2 之前的部分。
asyn	默认值为 False。该参数表示目标网站是否为异步加载。若为异步加载，则 asyn = True。若不为异步加载，则 asyn = False。
start_page	默认值为空值。该参数表示：想要爬取的起始页码（这里的页码不是经过数学运算的页码，而是你在页面中看到的），即，第几页。
end_page	默认值为空值。该参数表示：想要爬取的终止页码（这里的页码不是经过数学运算的页码，而是你在页面中看到的），即，第几页。
normal_page_number	<p>默认值为 True。该参数表示：url 中的页码是否存在数学运算。</p> <p>若存在数学运算，请将值修改为 False</p> <p>如：百度贴吧 url 中的页码：第一页 1 或 0，第二页 50，第三页 100。</p> <p>https://tieba.baidu.com/f?kw=pubg&ie=utf-8&pn=50</p> <p>若 url 页码是经过数学运算的，需修改参数：normal_page_number = False,</p> <p>my_requests.auto_requests(folder_name = XXX, url = XXX, normal_page_number = False)</p>
url_numbers	<p>默认值为空值。该参数表示：以列表或生成器形式表示的经数学运算后的 url 中的页码。</p> <p>若 url 页码需要数学运算，请将数学运算后的页码以列表或生成器形式传入模块。示例如下：</p> <p>百度贴吧 url：https://tieba.baidu.com/f?kw=pubg&ie=utf-8&pn=50 第一页为 1 或 0，第二页为 50，第三页为 100。</p> <p>列表形式：</p> <pre>x = [(i-1) * 50 for i in range(start_page, end_page + 1)] my_requests.auto_requests(folder_name = XXX, url = XXX, normal_page_number = False, url_numbers = x)</pre>

生成器形式：

`x = ((i-1) * 50 for i in range(start_page, end_page + 1))` 注意：生成器形式最外层是圆括号，列表形式最外层是方括号。

`my_requests.auto_requests(folder_name = XXX, url = XXX, normal_page_number = False, url_numbers = x)`

`url_rear`

默认值为 ' / '。该参数表示：url 后半部分，即，页码数字之后的部分。默认为"/"，若不为“/”，则可根据需求设定。

如人民网网址：`it.people.com.cn/index2.html`

其中：`index` 后的数字‘2’表示页码，页码后的部分为：‘.html’。故，这里，需设定：`url_rear = '.html'`。

完整代码：

```
url = 'it.people.com.cn/index2'
rear = '.html'
my_requests.auto_requests(folder_name = XXX, url = url, url_rear = rear)
```

`nap`

默认值 = 0。该参数表示单页休眠时间（单位：秒）： 即，若需下载多页，每页之间需间隔几秒后再进行下一页的下载。

若需添加每页下载后休眠 30 秒：

`my_requests.auto_requests(folder_name = XXX, url = XXX, nap = 30)`

`sleep`

默认值 = 0 。该参数表示批次休眠时间（单位：秒）： 若需多次下载某些页面，该参数表示，每完整下载完一个任务后需要间隔多久再进行下一次下载。

示例如下：

若想每 5 分钟下载一次 1 至 5 页，则 `sleep = 300`，即 `5*60=300`

`my_requests.auto_requests(folder_name = XXX, url = XXX, sleep = 300)`

注意：用户若想循环多次下载网页必须将 `sleep` 修改为非零值，此程序禁止用户连续无休止爬取网页。

ini_count	<p>默认值 = 0。该参数为子目录文件夹（sub_folder）的编号数字。</p> <p>此模块下载目录结构如下：</p> <p>folder_name:</p> <div><div>sub_folder_name_1:</div><div><div>file_1</div><div>file_2</div></div></div> <div><div>sub_folder_name_2:</div><div><div>file_1</div><div>file_2</div></div></div>
-----------	---

备注：带星号（*）参数为必填参数，其余参数为选填参数。

教程：

该模块所有参数一览：

```
auto_requests(folder_name, url, async=False, start_page = '', end_page = '', normal_page_number = True, url_numbers='', url_rear = '/',
               nap = 0, sleep = 0, ini_count = 0, single_folder = True, file_type = '.html')
```

单页下载示例：（该页面下载一次后程序结束）

```
from my_module import my_requests

url = 'https://coinmarketcap.com/'
my_requests.auto_requests('coinmktcap',url)
```

单页多次下载示例：（程序将每隔 30 秒下载一次 url 对应的网页）

```
from my_module import my_requests

#全部存放在一个同一根目录下方法：
url = 'https://coinmarketcap.com/'
my_requests.auto_requests('coinmktcap',url,sleep = 30)
#创建以时间戳命名的子目录，并将下载文件分别放置于子目录中的方法：
my_requests.auto_requests('coinmktcap',url,sleep = 30,single_folder = False)
```

常规页面多页多次下载示例：（程序将每隔 30 秒下载一次 1-3 页，每一页之间并无间歇）

```
from my_module import my_requests

url = 'https://coinmarketcap.com/'
my_requests.auto_requests('coinmktcap', url, start_page = 1, end_page = 3, sleep = 30,single_folder = False)
```

非常规 url 页码编号网页的多页多次下载示例：

```
from my_module import my_requests
```

```
url = 'https://tieba.baidu.com/f?kw=pubg&ie=utf-8&pn='
```

```
start_page = 1
```

```
end_page =3
```

```
x = ((i-1) * 50 for i in range(start_page, end_page + 1))
```

```
#上面使用的是生成器方式。若用户想以列表形式传入 url 中的页码: x = [i * 50 for i in range(start_page, end_page + 1)]
```

```
my_requests.auto_requests('tieba', url, start_page = 1, end_page = 3, normal_page_number = False, url_numbers = x, sleep = 10,single_folder = False)
```

异步加载网站的下载示例：（用户无需手动在浏览器中查找 json 文件，只需将 asyn 参数设置为 True 即可。）

```
from my_module import my_requests
```

```
url = 'https://tieba.baidu.com/f?kw=pubg&ie=utf-8&pn='
```

```
start_page = 1
```

```
end_page =3
```

```
x = ((i-1) * 50 for i in range(start_page, end_page + 1))
```

```
my_requests.auto_requests('tieba', url, asyn = True, start_page = 1, end_page = 3, normal_page_number = False, url_numbers = x, sleep = 10,single_folder = False)
```