用户指南

这部分文档是以文字为主，从 Requests 的背景讲起，然后对 Requests 的重点功能做了逐一的介绍。

简介

开发哲学

Requests 是以 **PEP 20** 的箴言为中心开发的

1. Beautiful is better than ugly.(美丽优于丑陋)
2. Explicit is better than implicit.(直白优于含蓄)
3. Simple is better than complex.(简单优于复杂)
4. Complex is better than complicated.(复杂优于繁琐)
5. Readability counts.(可读性很重要)

对于 Requests 所有的贡献都应牢记这些重要的准则。

Apache2 协议

现在你找到的许多开源项目都是以 GPL 协议发布的。虽然 GPL 有它自己的一席之地， 但在开始你的下一个开源项目时，GPL 应该不再是你的默认选择。

项目发行于 GPL 协议之后，就不能用于任何本身没开源的商业产品中。

MIT、BSD、ISC、Apache2 许可都是优秀的替代品，它们允许你的开源软件自由应用在私有闭源软件中。

Requests 的发布许可为 Apache2 License.

Requests 协议

安装

这部分文档包含了 Requests 的安装过程，使用任何软件的第一步就是正确地安装它。

Pip Install Requests

要安装 Requests，只要在你的终端中运行这个简单命令即可：

```
$ pip install requests
```

如果你没有安装 pip （啧啧），这个 Python installation guide 可以带你完成这一流程。

获得源码

Requests 一直在 Github 上积极地开发，你可以一直从这里获取到代码。

你可以克隆公共版本库：

```
git clone git://github.com/kennethreitz/requests.git
```

也可以下载 tarball:

```
$ curl -OL
https://github.com/kennethreitz/requests/tarball/master

  # Windows 用户也可选择 zip 包
```

获得代码之后，你就可以轻松的将它嵌入到你的 python 包里，或者安装到你的 site-packages:

```
$ python setup.py install
```

快速上手

迫不及待了吗？本页内容为如何入门 Requests 提供了很好的指引。其假设你已经安装了 Requests。如果还没有，去安装一节看看吧。

首先，确认一下：

- Requests 已安装
- Requests 是最新的

让我们从一些简单的示例开始吧。

发送请求

使用 Requests 发送网络请求非常简单。

一开始要导入 Requests 模块：

```
>>> import requests
```

然后，尝试获取某个网页。本例子中，我们来获取 Github 的公共时间线：

```
>>> r = requests.get('https://github.com/timeline.json')
```

现在，我们有一个名为 `r` 的 `Response` 对象。我们可以从这个对象中获取所有我们想要的信息。

Requests 简便的 API 意味着所有 HTTP 请求类型都是显而易见的。例如，你可以这样发送一个 HTTP POST 请求：

```
>>> r = requests.post("http://httpbin.org/post")
```

漂亮，对吧？那么其他 HTTP 请求类型：PUT，DELETE，HEAD 以及 OPTIONS 又是如何的呢？都是一样的简单：

```
>>> r = requests.put("http://httpbin.org/put")

>>> r = requests.delete("http://httpbin.org/delete")

>>> r = requests.head("http://httpbin.org/get")
```

```
>>> r = requests.options("http://httpbin.org/get")
```

都很不错吧，但这也仅是 Requests 的冰山一角呢。

传递 URL 参数

你也许经常想为 URL 的查询字符串(query string)传递某种数据。如果你是手工构建 URL，那么数据会以键/值对的形式置于 URL 中，跟在一个问号的后面。例如， `httpbin.org/get?key=val`。 Requests 允许你使用 `params` 关键字参数，以一个字典来提供这些参数。举例来说，如果你想传递 `key1=value1` 和 `key2=value2` 到 `httpbin.org/get` ，那么你可以使用如下代码：

```
>>> payload = {'key1': 'value1', 'key2': 'value2'}

>>> r = requests.get("http://httpbin.org/get", params=payload)
```

通过打印输出该 URL，你能看到 URL 已被正确编码：

```
>>> print(r.url)

http://httpbin.org/get?key2=value2&key1=value1
```

注意字典里值为 `None` 的键都不会被添加到 URL 的查询字符串里。

你还可以将一个列表作为值传入：

```
>>> payload = {'key1': 'value1', 'key2': ['value2', 'value3']}



>>> r = requests.get('http://httpbin.org/get', params=payload)

>>> print(r.url)

http://httpbin.org/get?key1=value1&key2=value2&key2=value3
```

响应内容

我们能读取服务器响应的内容。再次以 GitHub 时间线为例：

```
>>> import requests

>>> r = requests.get('https://github.com/timeline.json')
```

```
>>> r.text

u'[{"repository":{"open_issues":0,"url":"https://github.com/...
```

Requests 会自动解码来自服务器的内容。大多数 unicode 字符集都能被无缝地解码。

请求发出后，Requests 会基于 HTTP 头部对响应的编码作出有根据的推测。当你访问 `r.text` 之时，Requests 会使用其推测的文本编码。你可以找出 Requests 使用了什么编码，并且能够使用 `r.encoding` 属性来改变它：

```
>>> r.encoding

'utf-8'

>>> r.encoding = 'ISO-8859-1'
```

如果你改变了编码，每当你访问 `r.text` ，Request 都将会使用 `r.encoding` 的新值。你可能希望在使用特殊逻辑计算出文本的编码的情况下来修改编码。比如 HTTP 和 XML 自身可以指定编码。这样的话，你应该使用 `r.content` 来找到编码，然后设置 `r.encoding` 为相应的编码。这样就能使用正确的编码解析 `r.text` 了。

在你需要的情况下，Requests 也可以使用定制的编码。如果你创建了自己的编码，并使用 `codecs` 模块进行注册，你就可以轻松地使用这个解码器名称作为 `r.encoding` 的值， 然后由 Requests 来为你处理编码。

二进制响应内容

你也能以字节的方式访问请求响应体，对于非文本请求：

```
>>> r.content

b'[{"repository":{"open_issues":0,"url":"https://github.com/...
```

Requests 会自动为你解码 `gzip` 和 `deflate` 传输编码的响应数据。

例如，以请求返回的二进制数据创建一张图片，你可以使用如下代码：

```
>>> from PIL import Image

>>> from io import BytesIO

>>> i = Image.open(BytesIO(r.content))
```

JSON 响应内容

Requests 中也有一个内置的 JSON 解码器，助你处理 JSON 数据：

```
>>> import requests



>>> r = requests.get('https://github.com/timeline.json')

>>> r.json()

[{u'repository': {u'open_issues': 0, u'url': 'https://github.com/...
```

如果 JSON 解码失败， `r.json` 就会抛出一个异常。例如，相应内容是 401 (Unauthorized)，尝试访问 `r.json` 将会抛出 `ValueError: No JSON object could be decoded` 异常。

原始响应内容

在罕见的情况下，你可能想获取来自服务器的原始套接字响应，那么你可以访问 `r.raw`。 如果你确实想这么干，那请你确保在初始请求中设置了 `stream=True`。具体你可以这么做：

```
>>> r = requests.get('https://github.com/timeline.json', stream=True)

>>> r.raw

<requests.packages.urllib3.response.HTTPResponse object at 0x101194810>

>>> r.raw.read(10)

'\x1f\x8b\x08\x00\x00\x00\x00\x00\x00\x03'
```

但一般情况下，你应该以下面的模式将文本流保存到文件：

```
with open(filename, 'wb') as fd:

    for chunk in r.iter_content(chunk_size):

        fd.write(chunk)
```

使用 `Response.iter_content` 将会处理大量你直接使用 `Response.raw` 不得不处理的。 当流下载时，上面是优先推荐的获取内容方式。

定制请求头

如果你想为请求添加 HTTP 头部，只要简单地传递一个 `dict` 给 `headers` 参数就可以了。

例如，在前一个示例中我们没有指定 content-type:

```
>>> url = 'https://api.github.com/some/endpoint'

>>> headers = {'user-agent': 'my-app/0.0.1'}


>>> r = requests.get(url, headers=headers)
```

注意：定制 header 的优先级低于某些特定的信息源，例如：

- 如果在 `.netrc` 中设置了用户认证信息，使用 *headers=* 设置的授权就不会生效。而如果设置了 `auth=` 参数，``.netrc`` 的设置就无效了。
- 如果被重定向到别的主机，授权 header 就会被删除。
- 代理授权 header 会被 URL 中提供的代理身份覆盖掉。
- 在我们能判断内容长度的情况下，header 的 Content-Length 会被改写。

更进一步讲，Requests 不会基于定制 header 的具体情况改变自己的行为。只不过在最后的请求中，所有的 header 信息都会被传递进去。

注意：所有的 header 值必须是 `string`、bytestring 或者 unicode。尽管传递 unicode header 也是允许的，但不建议这样做。

更加复杂的 POST 请求

通常，你想要发送一些编码为表单形式的数据——非常像一个 HTML 表单。要实现这个，只需简单地传递一个字典给 *data* 参数。你的数据字典在发出请求时会自动编码为表单形式：

```
>>> payload = {'key1': 'value1', 'key2': 'value2'}
```

```
>>> r = requests.post("http://httpbin.org/post", data=payload)
>>> print(r.text)
{
  ...
  "form": {
    "key2": "value2",
    "key1": "value1"
  },
  ...
}
```

很多时候你想要发送的数据并非编码为表单形式的。如果你传递一个 `string` 而不是一个 `dict`，那么数据会被直接发布出去。

例如，Github API v3 接受编码为 JSON 的 POST/PATCH 数据：

```
>>> import json

>>> url = 'https://api.github.com/some/endpoint'
>>> payload = {'some': 'data'}

>>> r = requests.post(url, data=json.dumps(payload))
```

此处除了可以自行对 `dict` 进行编码，你还可以使用 `json` 参数直接传递，然后它就会被自动编码。这是 2.4.2 版的新加功能：

```
>>> url = 'https://api.github.com/some/endpoint'
>>> payload = {'some': 'data'}

>>> r = requests.post(url, json=payload)
```

POST 一个多部分编码(Multipart-Encoded)的文件

Requests 使得上传多部分编码文件变得很简单：

```
>>> url = 'http://httpbin.org/post'

>>> files = {'file': open('report.xls', 'rb')}


>>> r = requests.post(url, files=files)

>>> r.text

{

  ...

  "files": {

    "file": "<censored...binary...data>"

  },

  ...

}
```

你可以显式地设置文件名，文件类型和请求头：

```
>>> url = 'http://httpbin.org/post'

>>> files = {'file': ('report.xls', open('report.xls', 'rb'),
'application/vnd.ms-excel', {'Expires': '0'})}

>>> r = requests.post(url, files=files)

>>> r.text

{

  ...

  "files": {

    "file": "<censored...binary...data>"

  },

  ...

}
```

如果你想，你也可以发送作为文件来接收的字符串：

```
>>> url = 'http://httpbin.org/post'

>>> files = {'file': ('report.csv',
'some,data,to,send\nanother,row,to,send\n')}



>>> r = requests.post(url, files=files)

>>> r.text

{

  ...

  "files": {

    "file": "some,data,to,send\\nanother,row,to,send\\n"

  },

  ...

}
```

如果你发送一个非常大的文件作为 `multipart/form-data` 请求，你可能希望将请求做成数据流。默认下 `requests` 不支持，但有个第三方包 `requests-toolbelt` 是支持的。你可以阅读 toolbelt 文档来了解使用方法。

在一个请求中发送多文件参考 高级用法 一节。

> 警告
>
> 我们强烈建议你用二进制模式(binary mode)打开文件。这是因为 Requests 可能会试图为你提供 `Content-Length` header，在它这样做的时候，这个值会被设为文件的字节数（*bytes*）。如果用文本模式(text mode)打开文件，就可能会发生错误。

响应状态码

我们可以检测响应状态码：

```
>>> r = requests.get('http://httpbin.org/get')

>>> r.status_code

200
```

为方便引用，Requests 还附带了一个内置的状态码查询对象：

```
>>> r.status_code == requests.codes.ok

True
```

如果发送了一个错误请求(一个 4XX 客户端错误，或者 5XX 服务器错误响应)，我们可以通过 `Response.raise_for_status()` 来抛出异常：

```
>>> bad_r = requests.get('http://httpbin.org/status/404')

>>> bad_r.status_code

404



>>> bad_r.raise_for_status()

Traceback (most recent call last):

  File "requests/models.py", line 832, in raise_for_status

    raise http_error

requests.exceptions.HTTPError: 404 Client Error
```

但是，由于我们的例子中 `r` 的 `status_code` 是 `200`，当我们调用 `raise_for_status()` 时，得到的是：

```
>>> r.raise_for_status()

None
```

一切都挺和谐哈。

响应头

我们可以查看以一个 Python 字典形式展示的服务器响应头：

```
>>> r.headers

{

    'content-encoding': 'gzip',

    'transfer-encoding': 'chunked',

    'connection': 'close',

    'server': 'nginx/1.0.4',

    'x-runtime': '148ms',

    'etag': '"e1ca502697e5c9317743dc078f67693f"',

    'content-type': 'application/json'

}
```

但是这个字典比较特殊：它是仅为 HTTP 头部而生的。根据 RFC 2616，HTTP 头部是大小写不敏感的。

因此，我们可以使用任意大写形式来访问这些响应头字段：

```
>>> r.headers['Content-Type']

'application/json'



>>> r.headers.get('content-type')

'application/json'
```

它还有一个特殊点，那就是服务器可以多次接受同一 header，每次都使用不同的值。但 Requests 会将它们合并，这样它们就可以用一个映射来表示出来，参见 RFC 7230:

A recipient MAY combine multiple header fields with the same field name into one "field-name: field-value" pair, without changing the semantics of the message, by appending each subsequent field value to the combined field value in order, separated by a comma.

接收者可以合并多个相同名称的 header 栏位，把它们合为一个 "field-name: field-value" 配对，将每个后续的栏位值依次追加到合并的栏位值中，用逗号隔开即可，这样做不会改变信息的语义。

Cookie

如果某个响应中包含一些 cookie，你可以快速访问它们：

```
>>> url = 'http://example.com/some/cookie/setting/url'
>>> r = requests.get(url)


>>> r.cookies['example_cookie_name']
'example_cookie_value'
```

要想发送你的 cookies 到服务器，可以使用 `cookies` 参数：

```
>>> url = 'http://httpbin.org/cookies'
>>> cookies = dict(cookies_are='working')


>>> r = requests.get(url, cookies=cookies)
>>> r.text
'{"cookies": {"cookies_are": "working"}}'
```

重定向与请求历史

默认情况下，除了 HEAD, Requests 会自动处理所有重定向。

可以使用响应对象的 `history` 方法来追踪重定向。

`Response.history` 是一个 **`Response`** 对象的列表，为了完成请求而创建了这些对象。这个对象列表按照从最老到最近的请求进行排序。

例如，Github 将所有的 HTTP 请求重定向到 HTTPS：

```
>>> r = requests.get('http://github.com')


>>> r.url
'https://github.com/'
```

```
>>> r.status_code
200

>>> r.history
[<Response [301]>]
```

如果你使用的是 GET、OPTIONS、POST、PUT、PATCH 或者 DELETE，那么你可以通过 `allow_redirects` 参数禁用重定向处理：

```
>>> r = requests.get('http://github.com', allow_redirects=False)
>>> r.status_code
301
>>> r.history
[]
```

如果你使用了 HEAD，你也可以启用重定向：

```
>>> r = requests.head('http://github.com', allow_redirects=True)
>>> r.url
'https://github.com/'
>>> r.history
[<Response [301]>]
```

超时

你可以告诉 requests 在经过以 `timeout` 参数设定的秒数时间之后停止等待响应：

```
>>> requests.get('http://github.com', timeout=0.001)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
```

```
requests.exceptions.Timeout: HTTPConnectionPool(host='github.com',
port=80): Request timed out. (timeout=0.001)
```

注意

`timeout` 仅对连接过程有效，与响应体的下载无关。 `timeout` 并不是整个下载响应的时间限制，而是如果服务器在 `timeout` 秒内没有应答，将会引发一个异常（更精确地说，是在 `timeout` 秒内没有从基础套接字上接收到任何字节的数据时）

错误与异常

遇到网络问题（如：DNS 查询失败、拒绝连接等）时，Requests 会抛出一个 **`ConnectionError`** 异常。

如果 HTTP 请求返回了不成功的状态码， **`Response.raise_for_status()`** 会抛出一个 **`HTTPError`** 异常。

若请求超时，则抛出一个 **`Timeout`** 异常。

若请求超过了设定的最大重定向次数，则会抛出一个 **`TooManyRedirects`** 异常。

所有 Requests 显式抛出的异常都继承自 **`requests.exceptions.RequestException`** 。

高级用法

本篇文档涵盖了 Requests 的一些高级特性。

会话对象

会话对象让你能够跨请求保持某些参数。它也会在同一个 Session 实例发出的所有请求之间保持 cookie， 期间使用 `urllib3` 的 connection pooling 功能。所以如果你向同一主机发送多个请求，底层的 TCP 连接将会被重用，从而带来显著的性能提升。 (参见 HTTP persistent connection).

会话对象具有主要的 Requests API 的所有方法。

我们来跨请求保持一些 cookie:

```
s = requests.Session()



s.get('http://httpbin.org/cookies/set/sessioncookie/123456789')

r = s.get("http://httpbin.org/cookies")



print(r.text)

# '{"cookies": {"sessioncookie": "123456789"}}'
```

会话也可用来为请求方法提供缺省数据。这是通过为会话对象的属性提供数据来实现的：

```
s = requests.Session()

s.auth = ('user', 'pass')

s.headers.update({'x-test': 'true'})



# both 'x-test' and 'x-test2' are sent

s.get('http://httpbin.org/headers', headers={'x-test2': 'true'})
```

任何你传递给请求方法的字典都会与已设置会话层数据合并。方法层的参数覆盖会话的参数。

不过需要注意，就算使用了会话，方法级别的参数也不会被跨请求保持。下面的例子只会和第一个请求发送 cookie ，而非第二个：

```
s = requests.Session()



r = s.get('http://httpbin.org/cookies', cookies={'from-my': 'browser'})

print(r.text)

# '{"cookies": {"from-my": "browser"}}'



r = s.get('http://httpbin.org/cookies')

print(r.text)

# '{"cookies": {}}'
```

如果你要手动为会话添加 cookie，就使用 Cookie utility 函数 来操纵 **Session.cookies**。

会话还可以用作前后文管理器：

```
with requests.Session() as s:

    s.get('http://httpbin.org/cookies/set/sessioncookie/123456789')
```

这样就能确保 `with` 区块退出后会话能被关闭，即使发生了异常也一样。

从字典参数中移除一个值

有时你会想省略字典参数中一些会话层的键。要做到这一点，你只需简单地在方法层参数中将那个键的值设置为 `None` ，那个键就会被自动省略掉。

包含在一个会话中的所有数据你都可以直接使用。学习更多细节请阅读 会话 API 文档。

请求与响应对象

任何时候调用 requests.*() 你都在做两件主要的事情。其一，你在构建一个 *Request* 对象， 该对象将被发送到某个服务器请求或查询一些资源。其二，一旦 `requests` 得到一个从 服务器返回的响应就会产生一个 `Response` 对象。该响应对象包含服务器返回的所有信息， 也包含你原来创建的 `Request` 对象。如下是一个简单的请求，从 Wikipedia 的服务器得到 一些非常重要的信息：

```
>>> r = requests.get('http://en.wikipedia.org/wiki/Monty_Python')
```

如果想访问服务器返回给我们的响应头部信息，可以这样做：

```
>>> r.headers

{'content-length': '56170', 'x-content-type-options': 'nosniff', 'x-
cache':

'HIT from cp1006.eqiad.wmnet, MISS from cp1010.eqiad.wmnet', 'content-
encoding':

'gzip', 'age': '3080', 'content-language': 'en', 'vary': 'Accept-
Encoding,Cookie',

'server': 'Apache', 'last-modified': 'Wed, 13 Jun 2012 01:33:50 GMT',

'connection': 'close', 'cache-control': 'private, s-maxage=0, max-
age=0,

must-revalidate', 'date': 'Thu, 14 Jun 2012 12:59:39 GMT', 'content-
type':

'text/html; charset=UTF-8', 'x-cache-lookup': 'HIT from
cp1006.eqiad.wmnet:3128,

MISS from cp1010.eqiad.wmnet:80'}
```

然而，如果想得到发送到服务器的请求的头部，我们可以简单地访问该请求，然后是该请求的头部：

```
>>> r.request.headers

{'Accept-Encoding': 'identity, deflate, compress, gzip',

'Accept': '*/*', 'User-Agent': 'python-requests/0.13.1'}
```

准备的请求 （Prepared Request）

当你从 API 或者会话调用中收到一个 `Response` 对象时，`request` 属性其实是使用了 `PreparedRequest`。有时在发送请求之前，你需要对 body 或者 header（或者别的什么东西）做一些额外处理，下面演示了一个简单的做法：

```
from requests import Request, Session


s = Session()

req = Request('GET', url,

    data=data,

    headers=header

)

prepped = req.prepare()


# do something with prepped.body

# do something with prepped.headers


resp = s.send(prepped,

    stream=stream,

    verify=verify,

    proxies=proxies,

    cert=cert,

    timeout=timeout

)


print(resp.status_code)
```

由于你没有对 `Request` 对象做什么特殊事情，你立即准备和修改了 `PreparedRequest` 对象，然后把它和别的参数一起发送到 `requests.*` 或者 `Session.*`。

然而，上述代码会失去 Requests **`Session`** 对象的一些优势， 尤其 **`Session`** 级别的状态，例如 cookie 就不会被应用到你的请求上去。要获取一个带有状态的 **`PreparedRequest`**， 请用 **`Session.prepare_request()`** 取代 **`Request.prepare()`** 的调用，如下所示：

```
from requests import Request, Session

s = Session()

req = Request('GET',  url,

    data=data

    headers=headers

)



prepped = s.prepare_request(req)



# do something with prepped.body

# do something with prepped.headers



resp = s.send(prepped,

    stream=stream,

    verify=verify,

    proxies=proxies,

    cert=cert,

    timeout=timeout

)

print(resp.status_code)
```

SSL 证书验证

Requests 可以为 HTTPS 请求验证 SSL 证书，就像 web 浏览器一样。要想检查某个主机的 SSL 证书，你可以使用 `verify` 参数:

```
>>> requests.get('https://kennethreitz.com', verify=True)

requests.exceptions.SSLError: hostname 'kennethreitz.com' doesn't match
either of '*.herokuapp.com', 'herokuapp.com'
```

在该域名上我没有设置 SSL，所以失败了。但 Github 设置了 SSL:

```
>>> requests.get('https://github.com', verify=True)

<Response [200]>
```

对于私有证书，你也可以传递一个 CA_BUNDLE 文件的路径给 `verify`。你也可以设置 `REQUEST_CA_BUNDLE` 环境变量。

如果你将 `verify` 设置为 False，Requests 也能忽略对 SSL 证书的验证。

```
>>> requests.get('https://kennethreitz.com', verify=False)

<Response [200]>
```

默认情况下， `verify` 是设置为 True 的。选项 `verify` 仅应用于主机证书。

你也可以指定一个本地证书用作客户端证书，可以是单个文件（包含密钥和证书）或一个包含两个文件路径的元组:

```
>>> requests.get('https://kennethreitz.com', cert=('/path/server.crt',
'/path/key'))

<Response [200]>
```

如果你指定了一个错误路径或一个无效的证书:

```
>>> requests.get('https://kennethreitz.com',
cert='/wrong_path/server.pem')

SSLError: [Errno 336265225] _ssl.c:347: error:140B0009:SSL
routines:SSL_CTX_use_PrivateKey_file:PEM lib
```

警告

本地证书的私有 key 必须是解密状态。目前，Requests 不支持使用加密的 key。

CA 证书

Requests 默认附带了一套它信任的根证书，来自于 Mozilla trust store。然而它们在每次 Requests 更新时才会更新。这意味着如果你固定使用某一版本的 Requests，你的证书有可能已经 太旧了。

从 Requests 2.4.0 版之后，如果系统中装了 certifi 包，Requests 会试图使用它里边的 证书。这样用户就可以在不修改代码的情况下更新他们的可信任证书。

为了安全起见，我们建议你经常更新 certifi！

响应体内容工作流

默认情况下，当你进行网络请求后，响应体会立即被下载。可以通过 `stream` 参数覆盖这个行为，推迟下载响应体直到访问 **`Response.content`** 属性：

```
tarball_url = 'https://github.com/kennethreitz/requests/tarball/master'

r = requests.get(tarball_url, stream=True)
```

此时仅有响应头被下载下来了，连接保持打开状态，因此允许我们根据条件获取内容：

```
if int(r.headers['content-length']) < TOO_LONG:

  content = r.content

  ...
```

你可以进一步使用 **`Response.iter_content`** 和 **`Response.iter_lines`** 方法来控制工作流，或者以 **`Response.raw`** 从底层 urllib3 的 **`urllib3.HTTPResponse <urllib3.response.HTTPResponse`** 读取。

如果你在请求中把 `stream` 设为 `True`，Requests 无法将连接释放回连接池，除非你 消耗了所有的数据，或者调用了 **`Response.close`**。 这样会带来连接效率

低下的问题。如果你发现你在使用 `stream=True` 的同时还在部分读取请求的 body（或者完全没有读取 body），那么你就应该考虑使用 `contextlib.closing` (文档)， 如下所示：

```
from contextlib import closing



with closing(requests.get('http://httpbin.org/get', stream=True)) as r:

    # 在此处理响应。
```

保持活动状态（持久连接）

好消息——归功于 urllib3，同一会话内的持久连接是完全自动处理的！同一会话内你发出的任何请求都会自动复用恰当的连接！

注意：只有所有的响应体数据被读取完毕连接才会被释放为连接池；所以确保将 `stream` 设置为 `False` 或读取 `Response` 对象的 `content` 属性。

流式上传

Requests 支持流式上传，这允许你发送大的数据流或文件而无需先把它们读入内存。要使用流式上传，仅需为你的请求体提供一个类文件对象即可：

```
with open('massive-body') as f:

    requests.post('http://some.url/streamed', data=f)
```

警告

我们强烈建议你用二进制模式（binary mode）打开文件。这是因为 requests 可能会为你提供 header 中的 `Content-Length`，在这种情况下该值会被设为文件的**字节数**。如果你用**文本模式**打开文件，就可能碰到错误。

块编码请求

对于出去和进来的请求，Requests 也支持分块传输编码。要发送一个块编码的请求，仅需为你的请求体提供一个生成器（或任意没有具体长度的迭代器）：

```
def gen():

    yield 'hi'

    yield 'there'



requests.post('http://some.url/chunked', data=gen())
```

对于分块的编码请求，我们最好使用 **Response.iter_content()** 对其数据进行迭代。在理想情况下，你的 request 会设置 `stream=True`，这样你就可以通过调用 `iter_content` 并将分块大小参数设为 `None`，从而进行分块的迭代。如果你要设置分块的最大体积，你可以把分块大小参数设为任意整数。

POST 多个分块编码的文件

你可以在一个请求中发送多个文件。例如，假设你要上传多个图像文件到一个 HTML 表单，使用一个多文件 field 叫做 "images":

```
<input type="file" name="images" multiple="true" required="true"/>
```

要实现，只要把文件设到一个元组的列表中，其中元组结构为 `(form_field_name, file_info)`:

```
>>> url = 'http://httpbin.org/post'

>>> multiple_files = [

       ('images', ('foo.png', open('foo.png', 'rb'), 'image/png')),

       ('images', ('bar.png', open('bar.png', 'rb'), 'image/png'))]

>>> r = requests.post(url, files=multiple_files)

>>> r.text

{

  ...
```

```
  'files': {'images': ' ....'}

  'Content-Type': 'multipart/form-data;
boundary=3131623adb2043caaeb5538cc7aa0b3a',

  ...

}
```

警告

我们强烈建议你用二进制模式（binary mode）打开文件。这是因为 requests 可能会为你提供 header 中的 `Content-Length`，在这种情况下该值会被设为文件的**字节数**。如果你用**文本模式**打开文件，就可能碰到错误。

事件挂钩

Requests 有一个钩子系统，你可以用来操控部分请求过程，或信号事件处理。

可用的钩子:

`response`:

从一个请求产生的响应

你可以通过传递一个 `{hook_name: callback_function}` 字典给 `hooks` 请求参数 为每个请求分配一个钩子函数：

```
hooks=dict(response=print_url)
```

`callback_function` 会接受一个数据块作为它的第一个参数。

```
def print_url(r, *args, **kwargs):

    print(r.url)
```

若执行你的回调函数期间发生错误，系统会给出一个警告。

若回调函数返回一个值，默认以该值替换传进来的数据。若函数未返回任何东西， 也没有什么其他的影响。

我们来在运行期间打印一些请求方法的参数：

```
>>> requests.get('http://httpbin.org', hooks=dict(response=print_url))

http://httpbin.org

<Response [200]>
```

自定义身份验证

Requests 允许你使用自己指定的身份验证机制。

任何传递给请求方法的 `auth` 参数的可调用对象，在请求发出之前都有机会修改请求。

自定义的身份验证机制是作为 `requests.auth.AuthBase` 的子类来实现的，也非常容易定义。Requests 在 `requests.auth` 中提供了两种常见的的身份验证方案： `HTTPBasicAuth` 和 `HTTPDigestAuth` 。

假设我们有一个 web 服务，仅在 `X-Pizza` 头被设置为一个密码值的情况下才会有响应。虽然这不太可能，但就以它为例好了。

```
from requests.auth import AuthBase

class PizzaAuth(AuthBase):

    """Attaches HTTP Pizza Authentication to the given Request
object."""

    def __init__(self, username):

        # setup any auth-related data here

        self.username = username


    def __call__(self, r):

        # modify and return the request

        r.headers['X-Pizza'] = self.username

        return r
```

然后就可以使用我们的 PizzaAuth 来进行网络请求:

```
>>> requests.get('http://pizzabin.org/admin',
auth=PizzaAuth('kenneth'))

<Response [200]>
```

流式请求

使用 **requests.Response.iter_lines()** 你可以很方便地对流式 API （例如 Twitter 的流式 API ） 进行迭代。简单地设置 `stream` 为 `True` 便可以使用 **iter_lines()** 对相应进行迭代:

```
import json

import requests

r = requests.get('http://httpbin.org/stream/20', stream=True)

for line in r.iter_lines():


    # filter out keep-alive new lines

    if line:

        print(json.loads(line))
```

警告

iter_lines() 不保证重进入时的安全性。多次调用该方法 会导致部分收到的数据丢失。如果你要在多处调用它，就应该使用生成的迭代器对象:

```
lines = r.iter_lines()

# Save the first line for later or just skip it

first_line = next(lines)

for line in lines:

    print(line)
```

代理

如果需要使用代理，你可以通过为任意请求方法提供 `proxies` 参数来配置单个请求:

```
import requests


proxies = {

  "http": "http://10.10.1.10:3128",

  "https": "http://10.10.1.10:1080",

}


requests.get("http://example.org", proxies=proxies)
```

你也可以通过环境变量 `HTTP_PROXY` 和 `HTTPS_PROXY` 来配置代理。

```
$ export HTTP_PROXY="http://10.10.1.10:3128"

$ export HTTPS_PROXY="http://10.10.1.10:1080"


$ python

>>> import requests

>>> requests.get("http://example.org")
```

若你的代理需要使用 HTTP Basic Auth，可以使用 *http://user:password@host/* 语法:

```
proxies = {

    "http": "http://user:pass@10.10.1.10:3128/",

}
```

要为某个特定的连接方式或者主机设置代理，使用 *scheme://hostname* 作为 key，它会针对指定的主机和连接方式进行匹配。

```
proxies = {'http://10.20.1.128': 'http://10.10.1.10:5323'}
```

注意，代理 URL 必须包含连接方式。

SOCKS

2.10.0 新版功能.

除了基本的 HTTP 代理，Request 还支持 SOCKS 协议的代理。这是一个可选功能，若要使用， 你需要安装第三方库。

你可以用 `pip` 获取依赖:

```
$ pip install requests[socks]
```

安装好依赖以后，使用 SOCKS 代理和使用 HTTP 代理一样简单：

```
proxies = {

    'http': 'socks5://user:pass@host:port',

    'https': 'socks5://user:pass@host:port'

}
```

合规性

Requests 符合所有相关的规范和 RFC，这样不会为用户造成不必要的困难。但这种对规范的考虑导致一些行为对于不熟悉相关规范的人来说看似有点奇怪。

编码方式

当你收到一个响应时，Requests 会猜测响应的编码方式，用于在你调用 **`Response.text`** 方法时对响应进行解码。Requests 首先在 HTTP 头部检测是否存在指定的编码方式，如果不存在，则会使用 charade 来尝试猜测编码方式。

只有当 HTTP 头部不存在明确指定的字符集，并且 `Content-Type` 头部字段包含 `text` 值之时， Requests 才不去猜测编码方式。在这种情况下， RFC 2616 指定默认字符集必须是 `ISO-8859-1` 。Requests 遵从这一规范。如果你需要一种不同的编码方式，你可以手动设置 **`Response.encoding`** 属性，或使用原始的 **`Response.content`**。

HTTP 动词

Requests 提供了几乎所有 HTTP 动词的功能：GET、OPTIONS、HEAD、POST、PUT、PATCH、DELETE。以下内容为使用 Requests 中的这些动词以及 Github API 提供了详细示例。

我将从最常使用的动词 GET 开始。HTTP GET 是一个幂等方法，从给定的 URL 返回一个资源。因而，当你试图从一个 web 位置获取数据之时，你应该使用这个动词。一个使用示例是尝试从 Github 上获取关于一个特定 commit 的信息。假设我们想获取 Requests 的 commit `a050faf` 的信息。我们可以这样去做：

```
>>> import requests

>>> r =
requests.get('https://api.github.com/repos/kennethreitz/requests/git/commits/a050faf084662f3a352dd1a941f2c7c9f886d4ad')
```

我们应该确认 GitHub 是否正确响应。如果正确响应，我们想弄清响应内容是什么类型的。像这样去做：

```
>>> if (r.status_code == requests.codes.ok):

...     print r.headers['content-type']

...

application/json; charset=utf-8
```

可见，GitHub 返回了 JSON 数据，非常好，这样就可以使用 `r.json` 方法把这个返回的数据解析成 Python 对象。

```
>>> commit_data = r.json()



>>> print commit_data.keys()

[u'committer', u'author', u'url', u'tree', u'sha', u'parents',
u'message']

>>> print commit_data[u'committer']

{u'date': u'2012-05-10T11:10:50-07:00', u'email':
u'me@kennethreitz.com', u'name': u'Kenneth Reitz'}
```

```
>>> print commit_data[u'message']

makin' history
```

到目前为止，一切都非常简单。嗯，我们来研究一下 GitHub 的 API。我们可以去看看文档，但如果使用 Requests 来研究也许会更有意思一点。我们可以借助 Requests 的 OPTIONS 动词来看看我们刚使用过的 url 支持哪些 HTTP 方法。

```
>>> verbs = requests.options(r.url)

>>> verbs.status_code

500
```

额，这是怎么回事？毫无帮助嘛！原来 GitHub，与许多 API 提供方一样，实际上并未实现 OPTIONS 方法。这是一个恼人的疏忽，但没关系，那我们可以使用枯燥的文档。然而，如果 GitHub 正确实现了 OPTIONS，那么服务器应该在响应头中返回允许用户使用的 HTTP 方法，例如：

```
>>> verbs = requests.options('http://a-good-website.com/api/cats')

>>> print verbs.headers['allow']

GET,HEAD,POST,OPTIONS
```

转而去查看文档，我们看到对于提交信息，另一个允许的方法是 POST，它会创建一个新的提交。由于我们正在使用 Requests 代码库，我们应尽可能避免对它发送笨拙的 POST。作为替代，我们来玩玩 GitHub 的 Issue 特性。

本篇文档是回应 Issue #482 而添加的。鉴于该问题已经存在，我们就以它为例。先获取它。

```
>>> r =
requests.get('https://api.github.com/repos/kennethreitz/requests/issues
/482')

>>> r.status_code

200

>>> issue = json.loads(r.text)

>>> print(issue[u'title'])

Feature any http verb in docs
```

```
>>> print(issue[u'comments'])

3
```

Cool，有 3 个评论。我们来看一下最后一个评论。

```
>>> r = requests.get(r.url + u'/comments')

>>> r.status_code

200

>>> comments = r.json()

>>> print comments[0].keys()

[u'body', u'url', u'created_at', u'updated_at', u'user', u'id']

>>> print comments[2][u'body']

Probably in the "advanced" section
```

嗯，那看起来似乎是个愚蠢之处。我们发表个评论来告诉这个评论者他自己的愚蠢。那么，这个评论者是谁呢？

```
>>> print comments[2][u'user'][u'login']

kennethreitz
```

好，我们来告诉这个叫 Kenneth 的家伙，这个例子应该放在快速上手指南中。根据 GitHub API 文档，其方法是 POST 到该话题。我们来试试看。

```
>>> body = json.dumps({u"body": u"Sounds great! I'll get right on
it!"})

>>> url =
u"https://api.github.com/repos/kennethreitz/requests/issues/482/comment
s"



>>> r = requests.post(url=url, data=body)

>>> r.status_code

404
```

额，这有点古怪哈。可能我们需要验证身份。那就有点纠结了，对吧？不对。Requests 简化了多种身份验证形式的使用，包括非常常见的 Basic Auth。

```
>>> from requests.auth import HTTPBasicAuth

>>> auth = HTTPBasicAuth('fake@example.com', 'not_a_real_password')



>>> r = requests.post(url=url, data=body, auth=auth)

>>> r.status_code

201



>>> content = r.json()

>>> print(content[u'body'])

Sounds great! I'll get right on it.
```

太棒了！噢，不！我原本是想说等我一会，因为我得去喂我的猫。如果我能够编辑这条评论那就好了！幸运的是，GitHub 允许我们使用另一个 HTTP 动词 PATCH 来编辑评论。我们来试试。

```
>>> print(content[u"id"])

5804413



>>> body = json.dumps({u"body": u"Sounds great! I'll get right on it
once I feed my cat."})

>>> url =
u"https://api.github.com/repos/kennethreitz/requests/issues/comments/58
04413"



>>> r = requests.patch(url=url, data=body, auth=auth)

>>> r.status_code

200
```

非常好。现在，我们来折磨一下这个叫 Kenneth 的家伙，我决定要让他急得团团转，也不告诉他是我在捣蛋。这意味着我想删除这条评论。GitHub 允许我们使用完全名副其实的 DELETE 方法来删除评论。我们来清除该评论。

```
>>> r = requests.delete(url=url, auth=auth)

>>> r.status_code

204

>>> r.headers['status']

'204 No Content'
```

很好。不见了。最后一件我想知道的事情是我已经使用了多少限额（ratelimit）。查查看，GitHub 在响应头部发送这个信息，因此不必下载整个网页，我将使用一个 HEAD 请求来获取响应头。

```
>>> r = requests.head(url=url, auth=auth)

>>> print r.headers

...

'x-ratelimit-remaining': '4995'

'x-ratelimit-limit': '5000'

...
```

很好。是时候写个 Python 程序以各种刺激的方式滥用 GitHub 的 API，还可以使用 4995 次呢。

响应头链接字段

许多 HTTP API 都有响应头链接字段的特性，它们使得 API 能够更好地自我描述和自我显露。

GitHub 在 API 中为 分页 使用这些特性，例如:

```
>>> url =
'https://api.github.com/users/kennethreitz/repos?page=1&per_page=10'

>>> r = requests.head(url=url)

>>> r.headers['link']
```

```
'<https://api.github.com/users/kennethreitz/repos?page=2&per_page=10>;
rel="next",
<https://api.github.com/users/kennethreitz/repos?page=6&per_page=10>;
rel="last"'
```

Requests 会自动解析这些响应头链接字段，并使得它们非常易于使用:

```
>>> r.links["next"]

{'url':
'https://api.github.com/users/kennethreitz/repos?page=2&per_page=10',
'rel': 'next'}



>>> r.links["last"]

{'url':
'https://api.github.com/users/kennethreitz/repos?page=7&per_page=10',
'rel': 'last'}
```

传输适配器

从 v1.0.0 以后，Requests 的内部采用了模块化设计。部分原因是为了实现传输适配器（Transport Adapter），你可以看看关于它的最早描述。传输适配器提供了一个机制，让你可以为 HTTP 服务定义交互方法。尤其是它允许你应用服务前的配置。

Requests 自带了一个传输适配器，也就是 **`HTTPAdapter`**。 这个适配器使用了强大的 urllib3，为 Requests 提供了默认的 HTTP 和 HTTPS 交互。每当 **`Session`** 被初始化，就会有适配器附着在 **`Session`** 上，其中一个供 HTTP 使用，另一个供 HTTPS 使用。

Request 允许用户创建和使用他们自己的传输适配器，实现他们需要的特殊功能。创建好以后，传输适配器可以被加载到一个会话对象上，附带着一个说明，告诉会话适配器应该应用在哪个 web 服务上。

```
>>> s = requests.Session()

>>> s.mount('http://www.github.com', MyAdapter())
```

这个 mount 调用会注册一个传输适配器的特定实例到一个前缀上面。加载以后，任何使用该会话的 HTTP 请求，只要其 URL 是以给定的前缀开头，该传输适配器就会被使用到。

传输适配器的众多实现细节不在本文档的覆盖范围内，不过你可以看看接下来这个简单的 SSL 用例。更多的用法，你也许该考虑为 ``requests.adapters.BaseAdapter`` 创建子类。

示例：指定的 SSL 版本

Requests 开发团队刻意指定了内部库（urllib3）的默认 SSL 版本。一般情况下这样做没有问题，不过是不是你可能会需要连接到一个服务节点，而该节点使用了和默认不同的 SSL 版本。

你可以使用传输适配器解决这个问题，通过利用 HTTPAdapter 现有的大部分实现，再加上一个 *ssl_version* 参数并将它传递到 `urllib3` 中。我们会创建一个传输适配器，用来告诉 `urllib3` 让它使用 SSLv3：

```
import ssl


from requests.adapters import HTTPAdapter

from requests.packages.urllib3.poolmanager import PoolManager




class Ssl3HttpAdapter(HTTPAdapter):

    """"Transport adapter" that allows us to use SSLv3."""


    def init_poolmanager(self, connections, maxsize, block=False):

        self.poolmanager = PoolManager(num_pools=connections,

                                       maxsize=maxsize,

                                       block=block,

                                       ssl_version=ssl.PROTOCOL_SSLv3)
```

阻塞和非阻塞

使用默认的传输适配器，Requests 不提供任何形式的非阻塞 IO。 `Response.content` 属性会阻塞，直到整个响应下载完成。如果你需要更多精细控制，该库的数据流功能（见 流式请求） 允许你每次接受少量的一部分响应，不过这些调用依然是阻塞式的。

如果你对于阻塞式 IO 有所顾虑，还有很多项目可以供你使用，它们结合了 Requests 和 Python 的某个异步框架。典型的优秀例子是 grequests 和 requests-futures。

Header 排序

在某些特殊情况下你也许需要按照次序来提供 header，如果你向 `headers` 关键字参数传入一个 `OrderedDict`，就可以向提供一个带排序的 header。**然而**，Requests 使用的默认 header 的次序会被优先选择，这意味着如果你在 `headers` 关键字参数中覆盖了默认 header，和关键字参数中别的 header 相比，它们也许看上去会是次序错误的。

如果这个对你来说是个问题，那么用户应该考虑在 `Session` 对象上面设置默认 header，只要将 `Session` 设为一个定制的 `OrderedDict` 即可。这样就会让它成为优选的次序。

超时（timeout）

为防止服务器不能及时响应，大部分发至外部服务器的请求都应该带着 timeout 参数。如果没有 timeout，你的代码可能会挂起若干分钟甚至更长时间。

连接超时指的是在你的客户端实现到远端机器端口的连接时（对应的是 `connect()`_），Request 会等待的秒数。一个很好的实践方法是把连接超时设为比 3 的倍数略大的一个数值，因为 TCP 数据包重传窗口 (TCP packet retransmission window) 的默认大小是 3。

一旦你的客户端连接到了服务器并且发送了 HTTP 请求，**读取**超时指的就是客户端等待服务器发送请求的时间。（特定地，它指的是客户端要等待服务器发送字节**之间**的时间。在 99.9% 的情况下这指的是服务器发送第一个字节之前的时间）。

如果你制订了一个单一的值作为 timeout，如下所示：

```
r = requests.get('https://github.com', timeout=5)
```

这一 timeout 值将会用作 `connect` 和 `read` 二者的 timeout。如果要分别制定，就传入一个元组：

```
r = requests.get('https://github.com', timeout=(3.05, 27))
```

如果远端服务器很慢，你可以让 Request 永远等待，传入一个 None 作为 timeout 值，然后就冲咖啡去吧。

```
r = requests.get('https://github.com', timeout=None)
```

身份认证

本篇文档讨论如何配合 Requests 使用多种身份认证方式。

许多 web 服务都需要身份认证，并且也有多种不同的认证类型。 以下，我们会从简单到复杂概述 Requests 中可用的几种身份认证形式。

基本身份认证

许多要求身份认证的 web 服务都接受 HTTP Basic Auth。这是最简单的一种身份认证，并且 Requests 对这种认证方式的支持是直接开箱即可用。

以 HTTP Basic Auth 发送请求非常简单：

```
>>> from requests.auth import HTTPBasicAuth

>>> requests.get('https://api.github.com/user',
auth=HTTPBasicAuth('user', 'pass'))

<Response [200]>
```

事实上，HTTP Basic Auth 如此常见，Requests 就提供了一种简写的使用方式：

```
>>> requests.get('https://api.github.com/user', auth=('user', 'pass'))

<Response [200]>
```

像这样在一个元组中提供认证信息与前一个 `HTTPBasicAuth` 例子是完全相同的。

netrc 认证

如果认证方法没有收到 `auth` 参数，Requests 将试图从用户的 netrc 文件中获取 URL 的 hostname 需要的认证身份。

如果找到了 hostname 对应的身份，就会以 HTTP Basic Auth 的形式发送请求。

摘要式身份认证

另一种非常流行的 HTTP 身份认证形式是摘要式身份认证，Requests 对它的支持也是开箱即可用的：

```
>>> from requests.auth import HTTPDigestAuth

>>> url = 'http://httpbin.org/digest-auth/auth/user/pass'

>>> requests.get(url, auth=HTTPDigestAuth('user', 'pass'))

<Response [200]>
```

OAuth 1 认证

Oauth 是一种常见的 Web API 认证方式。`requests-oauthlib` 库可以让 Requests 用户简单地创建 OAuth 认证的请求：

::

```
>>> import requests

>>> from requests_oauthlib import OAuth1

>>> url =
'https://api.twitter.com/1.1/account/verify_credentials.json'

>>> auth = OAuth1('YOUR_APP_KEY', 'YOUR_APP_SECRET',

                  'USER_OAUTH_TOKEN', 'USER_OAUTH_TOKEN_SECRET')

>>> requests.get(url, auth=auth)

<Response [200]>
```

关于 OAuth 工作流程的更多信息，请参见 OAuth 官方网站。 关于 requests-oauthlib 的文档和用例，请参见 GitHub 的 requests_oauthlib 代码库。

其他身份认证形式

Requests 的设计允许其他形式的身份认证用简易的方式插入其中。开源社区的成员时常为更复杂或不那么常用的身份认证形式编写认证处理插件。其中一些最优秀的已被收集在 Requests organization 页面中，包括:

- Kerberos
- NTLM

如果你想使用其中任何一种身份认证形式，直接去它们的 GitHub 页面，依照说明进行。

新的身份认证形式

如果你找不到所需要的身份认证形式的一个良好实现，你也可以自己实现它。Requests 非常易于添加你自己的身份认证形式。

要想自己实现，就从 **requests.auth.AuthBase** 继承一个子类，并实现 `__call__()` 方法：

```
>>> import requests
>>> class MyAuth(requests.auth.AuthBase):
...     def __call__(self, r):
...         # Implement my authentication
...         return r
...
>>> url = 'http://httpbin.org/get'
>>> requests.get(url, auth=MyAuth())
<Response [200]>
```

当一个身份认证模块被附加到一个请求上，在设置 request 期间就会调用该模块。因此 `__call__`方法必须完成使得身份认证生效的所有事情。一些身份认证形式会额外地添加钩子来提供进一步的功能。

你可以在 Requests organization 页面的 `auth.py` 文件中找到更多示例。