



Live de Python #22

Requests / Web scraping #3





Roteiro

- O que são requests?
- Temos várias opções
- Porque escolhi requests?
- Uma olhada na sintaxe básica dos requests
- CODE !!!




O que são requests? (recap. da Live #9)

Sempre que nosso navegador se comunica com um servidor web a comunicação é feita usando o protocolo HTTP que funciona em uma estrutura solicitação-resposta (request-response) e requisições são feitas usando verbos e as respostas são respondidas por códigos.



O que são requests? (Requisições - Verbos)

Verbo	O que solicita?
GET	Pete um recurso ao servidor
HEAD	Obtém somente o cabeçalho do recurso (metadados)
POST	Envia dados para o servidor processar (criar um novo recurso)
PUT	Cria ou edita algum recurso
DELETE	Solicita a remoção de um recurso



O que são requests? (Códigos de retorno)

Família de retorno	O que quer dizer?
1xx	Envia informações para o cliente de que sua requisição foi recebida e está sendo processada
2xx	Indica que a requisição do cliente foi bem sucedida
3xx	Informa a ação adicional que deve ser tomada para completar a requisição
4xx	Avisa que o cliente fez uma requisição que não pode ser atendida
5xx	Ocorreu um erro no servidor ao cumprir uma requisição válida.



Porque escolhi requests?

Como todos sabem o Python tem bibliotecas próprias para fazer requisições web como a **urllib** (1, 2, 3). O fato de ter 3 opções e o request ser padrão, pra mim já diz muita coisa. Não tenho que aprender muita coisa quando as libs e mudam e a quantidade de código do requests comparado as urllibs são absurdas.



Porque escolhi requests?

“Requests é uma biblioteca HTTP elegante e simples para o Python, construída para seres humanos.”

Philosophy

Requests was developed with a few PEP 20 idioms in mind.

1. Beautiful is better than ugly.
2. Explicit is better than implicit.
3. Simple is better than complex.
4. Complex is better than complicated.
5. Readability counts.

All contributions to Requests should keep these important rules in mind.

Sintaxe básica da lib [o]



```
from requests import get
```

```
google = get('http://google.com')
```

```
google.status_code # 200
```

```
google.headers # Cabeçalho da resposta
```

```
google.cookies # cookies da requisição
```

```
google.text # Sim, o HTML que vamos usar com bs4
```


Sintaxe básica da lib [1]



```
from requests import post

google = post('http://google.com')

google.status_code # 405

google.reason # Método não permitido

google.encoding # UTF-8

google.close # encerra a comunicação
```

—

COOOOODEEEEEEEEEEE !!!!