



# Python e Web scraping

Aprendendo a extrair  
dados da Web

## **Aula 3 - Web scraping**

Felipe Andrade  
felipe@potelo.com.br



Cookies?

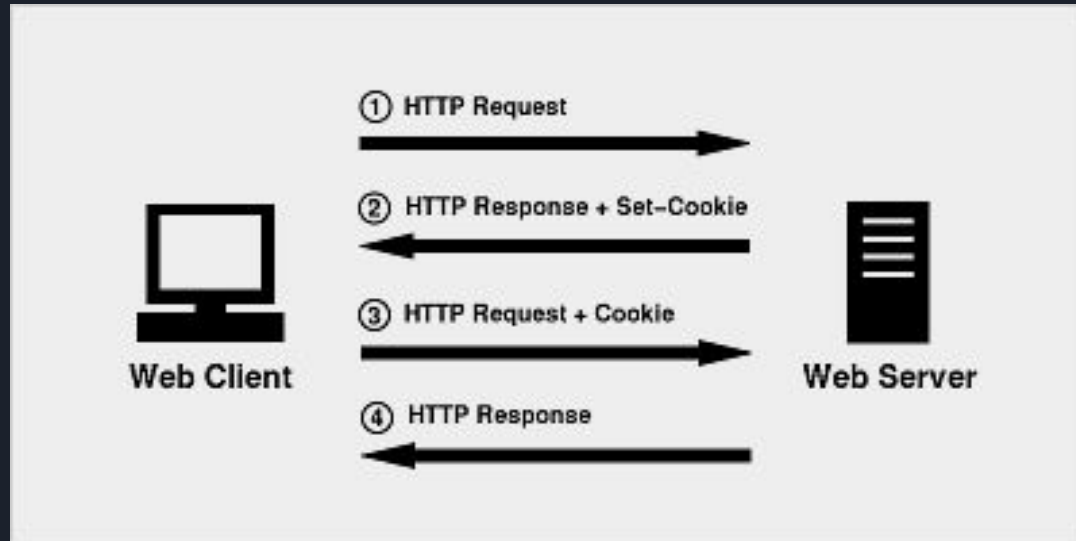




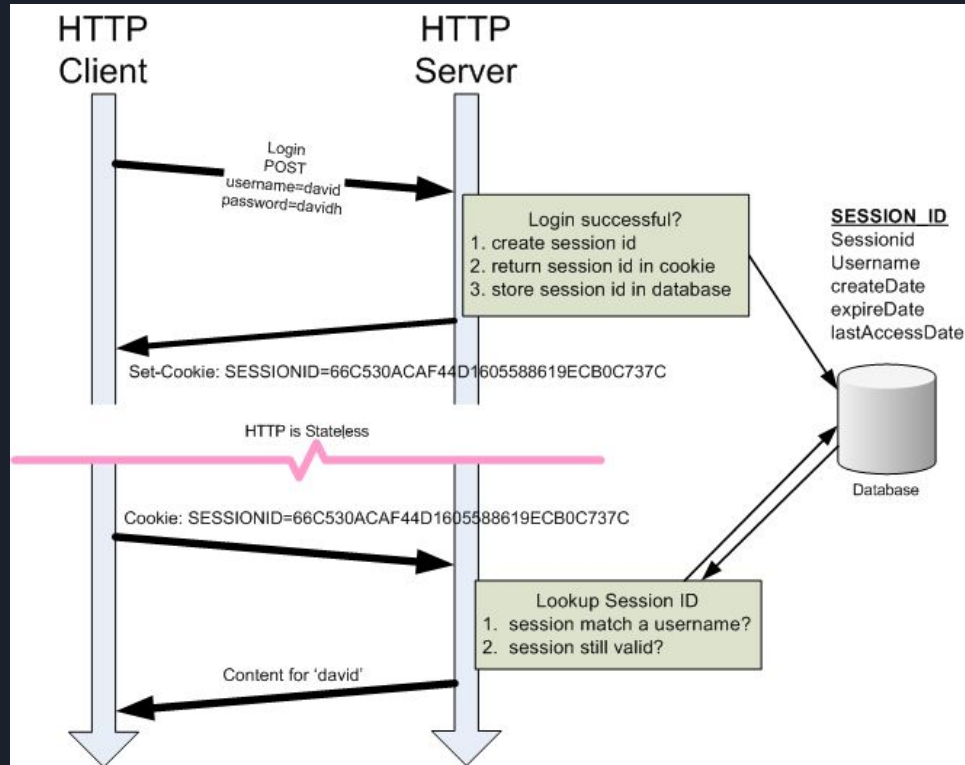
# HTTP - Cookies

- É um pequeno dado enviado pelo servidor no header do response;
- O browser armazena para quando for feita outra requisição;
- Podem ou não ter data de validade.

# HTTP - Cookies



# HTTP - Session



Hora de codar!





# Links interessantes

## Serviço de Proxy:

- <https://luminati.io> - 0.50 USD por IP por Mês

## Serviço de Captcha:

- <http://www.deathbycaptcha.com> - 6.95 USD por 5k de captcha
- <https://2captcha.com> - 0.5~1.0 USD por 1k de captcha

## Outras Libs similar ao Parsel:

- BeautifulSoup - <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- lxml - <http://lxml.de/>

## Similar ao Requests:

- Selenium - <http://selenium-python.readthedocs.io/>
- urllib - <https://docs.python.org/3/library/urllib.request.html>



# Livros interessantes

- Automatize tarefas maçantes com Python (Automate the Boring Stuff with Python)
  - Autor: Al Sweigart
- Web Scraping com Python (Web Scraping with Python)
  - Autor: Ryan Mitchell
- Web Scraping with Python
  - Autor: Richard Lawson





**Felipe Andrade**

felipe@potelo.com.br