

Mini Curso Python - Aula 3

Baseado nas aulas dos Profs. **Fernando Masanori, Álvaro Justen** (Turicas) e **Pedro Burgos**.

Josir Cardos Gomes / Doutorando em CI
Analista de Sistemas e Mestre em Administração

Roteiro

Aula 1: Estrutura da Linguagem

Aula 2: Lendo e gravando arquivos CSV

Aula 3: Lendo páginas na Web - Scrapy

- * Como funciona a Internet
- * Formatos de Arquivo
- * Scrapy

Aula 4: Acessando Twitter / Facebook / Scielo

Como funciona a Internet

Infra-Estrutura: Cabos e Hardware

Protocolos: IP - TCP - HTTP

Programas: Navegador

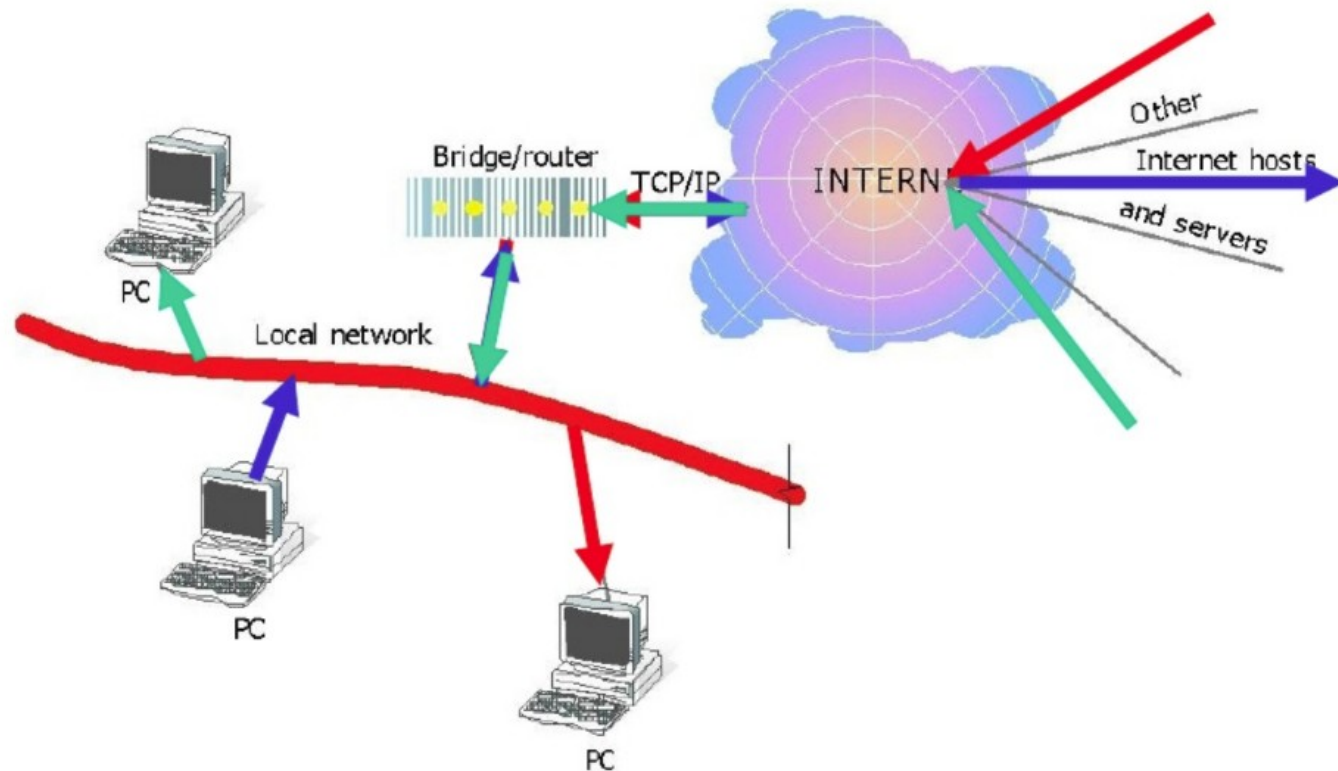
Chrome

Firefox

Safari

Internet Explorer

Python!



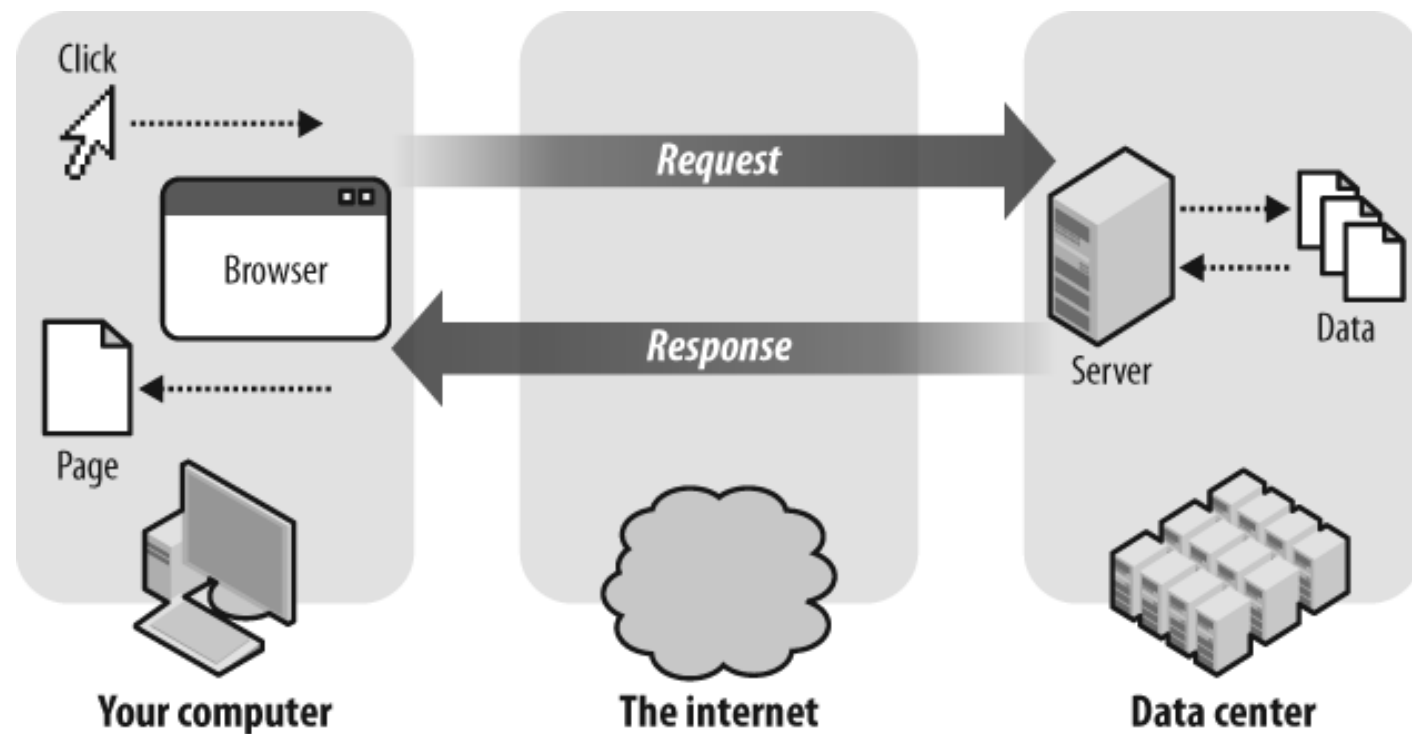
Endereço WEB

<http://www.farmi.pro.br/python/>

<http://> - Protocolo

www.farmi.pro.br - Site

[/python](http://www.farmi.pro.br/python/) - Diretório, Arquivo ou Serviço



Request → Response

O que vem como retorno ?

HTML, XML, JSON, TEXTO, uma imagem, um arquivo PDF, etc.

Como recuperar estes dados ?

* **Página Web (Scraping)**

* **Arquivo (CSV, XLS, etc)**

* **API**

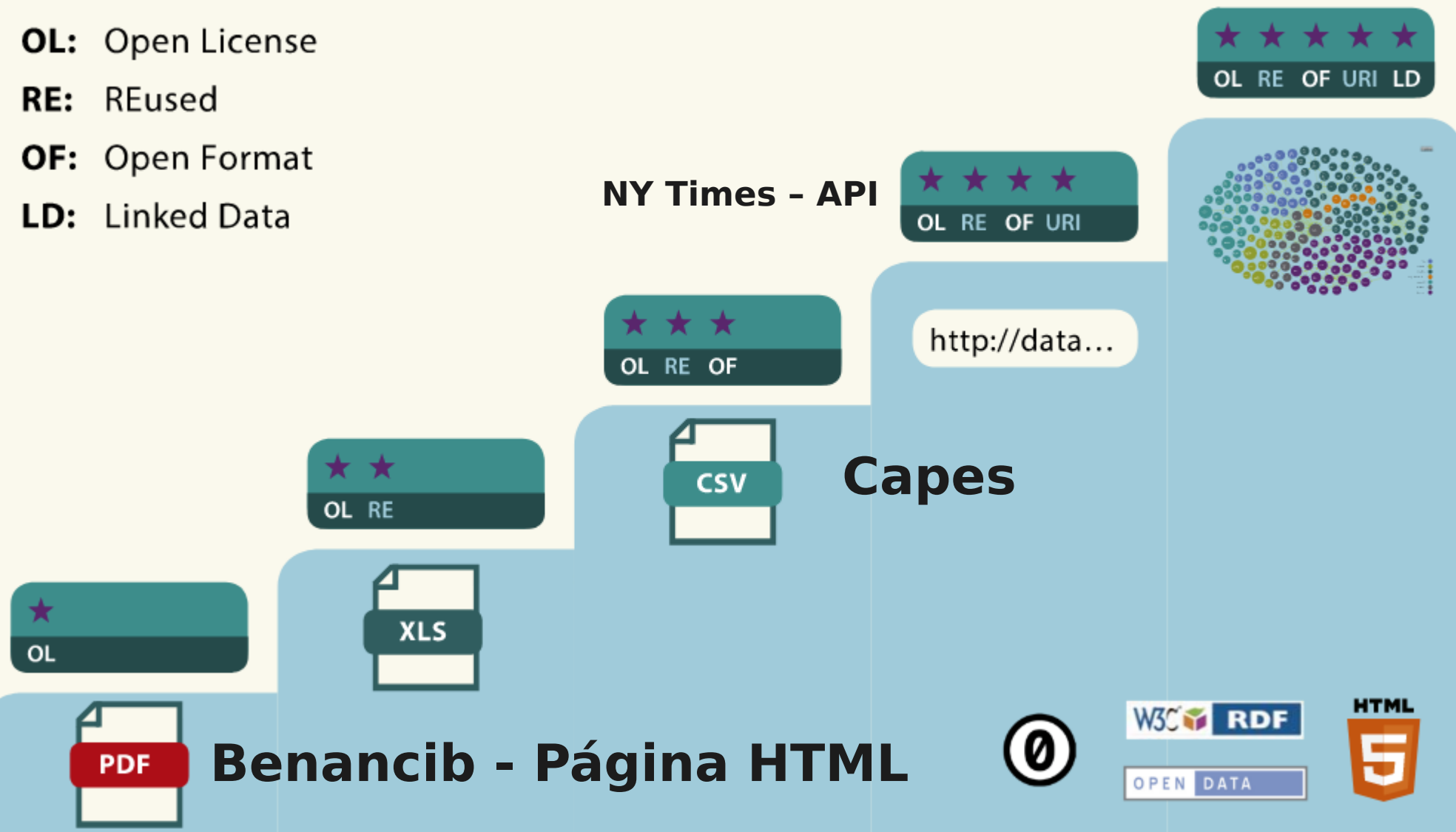
Roteiro e Resumo Geral

OL: Open License

RE: REused

OF: Open Format

LD: Linked Data



Scraping (ou Raspagem)

Obtem dados de sites que não disponibilizam APIs ou arquivos!

É um trabalho de detetive!

Fase Exploratória → Fase Planejamento → Fase Execução

Exemplo Inicial: <http://www.gilenofilho.com.br> Notebook: **scrap1.ipynb**

Google Chrome é nosso amigo! (F12)

O Safari também!

Safari > Preferences & then select the Advanced tab.

Select: "Show Develop menu in menu bar"

Click "Develop" in your menu bar and choose "Show Web Inspector"

HTML

```
<html>
<header>metadados</header>
<body>
  <a href="/hiperlink"></a>
  <div></div>
  <table>
    <tr> Linha
      <td></td> Coluna
    </tr>
  </table>
</body>
<scripts></scripts>
</html>
```


Scraping Avançado

Obtem dados de várias páginas!

Caso do BENANCIB

<http://repositorios.questoesemrede.uff.br/repositorios/advanced-search>

Notebook: **scrap2.ipynb**

Geralmente uma API requer um Login/Chave

- Limitar quem pode acessar os dados
- Restringir um número excessivo de requisições
- Registrar o que cada usuário está consultando ou gravando

Exemplo do NYTimes

Exemplo da Câmara dos Deputados

Exemplo Spotify

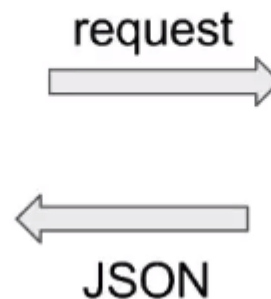
Facebook

Developers.facebook.com

Tools and Support (Ferramentas)

Explorador da Graph API

O que queremos?



Exercícios para próxima aula

Como listar o link do artigo na API do NyTimes ?

Existe uma API para a SCOPUS ou para WoS ?

Lembrando o link do roteiro do nosso curso:

<http://www.farmi.pro.br/python>