

Escavando dados com Python

Módulo Final



O que aprendemos?

- Git
- Sobre a linguagem Python
- Variáveis de dados em Python
- Condições e Repetições
- Estrutura de dados com python (Lista, dicionário, tupla)
- Funções (def)
- Princípios de orientação a objetos em Python

Parabéns!

- Se você conseguiu entender maioria dos tópicos citados, você já está no nível intermediário em Python

O que fazer agora?



Vamos aprimorar nosso bar!

- **Web scraping:** Extrair conteúdo de websites.

Habilidad Requerida



Instalar plugins

- **BeautifulSoup4 (bs4)**
- **python-requests**

Extrair Cardápio da internet



Você precisa estudar o site

- **Use o inspetor para entender a estrutura HTML do site**

O que descobrimos do site do cardápio?

- As tags <h3> são relacionadas a categorias
- As tags <h4> são relacionadas as sub-categorias
- As tags <h5> são relacionadas ao nome dos itens
- A especificação não possui tag HTML
- Infelizmente, o site não possui containers separados (container de bebidas, de comida e etc)
- **RESULTADO:** Difícil de escavar, mas ainda é possível!

Vamos decifrar o HTML na visão do BeautifulSoup

```
><h3 id="bebidas" class="azul">...</h3>


<h4 class="azul" id="caipirinhas">Caipirinhas</h4>
<h5>Sakerinha 350ml</h5>
"
    Morango"
<br>
<h5>Orloff Citrus 350ml</h5>
"
    Uva e Hortelã" == $0
<br>
<h5>Orloff Cranberry 350ml</h5>
"
    Morango e Cereja"
<br>
<h5>Orloff Green Apple 350ml</h5>
"
    Limão e Lima-da-Pérsia"
<br>
<h5>Orloff Natural 350ml</h5>
```

Procure por todas as tags <h4> (find_all)

Sibling (Tags que são irmãs [no mesmo nível HTML])

O que descobrimos de outro site de cardápio?

- **Cada panel representa a categoria e os items**
- **A classe pane-title são as categorias**
- **A classe list-group-item são os items**
- **RESULTADO: Fácil de escavar!**

O que descobrimos de outro site de cardápio?

[illegible]

Bônus

- **Criar rastreador de objetos do Correio!!!**

Enfim...

