

Automatizando o Download de Apostilas de Robótica da SEED-PR no Google Colab

Adriano Dubiel*

November 22, 2023

1 Montando o Google Drive no Colab

O código começa montando o Google Drive no ambiente do Google Colab. Isso permite o acesso e armazenamento de arquivos diretamente no Google Drive.

```
1 from google.colab import drive
2 drive.mount('/content/drive')
```

2 Obtendo URLs das Apostilas

A seguinte seção utiliza BeautifulSoup para fazer scraping da página que contém os links das apostilas em PDF. Os links são filtrados usando expressões regulares para garantir que sejam arquivos PDF.

```
1 from urllib.request import urlopen
2 from bs4 import BeautifulSoup
3 import re
4
5 html = urlopen("https://aluno.
6     escoladigital.pr.gov.br/
7     robotica/aulas/educacional")
8 bsObj = BeautifulSoup(html.read(),
9     'html.parser')
10 urls_unicas = set()
11
12 for link in bsObj.find_all('a', {
13     'href': re.compile('https://
14     aluno.escoladigital.pr.gov.br/
```

```
sites/alunos/arquivos_restritos
/files/documento/.*\.pdf'))):
    url = link.get('href')
    urls_unicas.add(url.strip())
```

3 Download e Salvamento no Google Drive

A próxima parte do código utiliza a biblioteca requests para realizar o download das apostilas. A barra de progresso é exibida usando a biblioteca tqdm para fornecer uma indicação visual do progresso do download.

```
1 import os
2 import requests
3 from tqdm import tqdm
4
5 drive_path = '/content/drive/
6     MyDrive/apostilas_robotica'
7 os.makedirs(drive_path, exist_ok=
8     True)
9
10 with tqdm(total=len(urls_unicas),
11     desc='Baixando apostilas', unit
12     ='apostilas') as pbar:
13     for url in urls_unicas:
14         response = requests.get(
15             url)
16
17         if response.status_code ==
18             200:
19             arquivo_path = os.path
20                 .join(drive_path, os.path.
21                     basename(url))
22             with open(arquivo_path
23                 , 'wb') as f:
24                 f.write(response.
25                     content)
```

*Professor de Física, Matemática e Robótica.¹⁴
Mestre em Física. Colégio Estadual do Campo Duque
de Caxias, Antônio Olinto, Paraná, NRE de União da¹⁵
Vitória. E-mail: adriano.dubiel@escola.pr.gov.br

```
16
17 pbar.update(1)
```

4 Agradecimentos

Gostaria de expressar meus sinceros agradecimentos ao Marcelo Santos, da Universidade Federal do Paraná, por sua inestimável ajuda e orientação ao longo do desenvolvimento deste projeto.

5 Conclusão

O código conclui imprimindo uma mensagem informando o número de apostilas de robótica baixadas e salvas no Google Drive.

```
1 print('O programa baixou', len(
    urls_unicas), 'apostilas de
    robótica da SEED/PR do ano de
    2023 e as salvou no Google
    Drive.')
```

6 Considerações Finais

O código apresentado é uma ferramenta eficaz para coletar e organizar material educacional de forma automatizada, facilitando o acesso a apostilas de Robótica da SEED-PR. Adriano Dubiel, o autor deste código, demonstra uma abordagem prática para a automação de tarefas relacionadas ao download e organização de recursos educacionais.

7 Referências

- Adriano Dubiel - GitHub: <https://github.com/ProfessorAdrianoDubiel/Robotica-Parana-2023>
- Marcelo Santos - Universidade Federal do Paraná
- Overleaf: <https://www.overleaf.com>