Automating the Download of Robotics Manuals from SEED-PR on Google Colab

Adriano Dubiel*

November 22, 2023

1 Mounting Google Drive on Colab

The code begins by mounting Google Drive in the Google Colab environment, enabling access and storage of files directly on Google Drive.

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

2 Obtaining Manual URLs

The following section uses BeautifulSoup for scraping the page containing links to robotics manuals in PDF format. The links are filtered using regular expressions to ensure they are PDF files.

```
from urllib.request import urlopen
from bs4 import BeautifulSoup
import re

html = urlopen("https://aluno.escoladigital.pr.gov.br/robotica/aulas/educacional")
bsObj = BeautifulSoup(html.read(), 'html.parser')
unique_urls = set()

for link in bsObj.find_all('a', {'href': re.compile('https://aluno.escoladigital.pr.gov.br/sites/alunos/arquivos_restritos/files/documento/.*\.pdf')}):
    url = link.get('href')
    unique_urls.add(url.strip())
```

3 Downloading and Saving to Google Drive

The next part of the code uses the requests library to download the manuals. A progress bar is displayed using the tqdm library to provide a visual indication of the download progress.

```
import os
import requests
from tqdm import tqdm

drive_path = '/content/drive/MyDrive/robotics_manuals'
os.makedirs(drive_path, exist_ok=True)

with tqdm(total=len(unique_urls), desc='Downloading manuals', unit='manuals') as pbar:
    for url in unique_urls:
        response = requests.get(url)

        if response.status_code == 200:
            file_path = os.path.join(drive_path, os.path.basename(url))
            with open(file_path, 'wb') as f:
                f.write(response.content)

        pbar.update(1)
```

4 Acknowledgments

I would like to express my sincere thanks to Marcelo Santos from the Federal University of

*Physics, Mathematics, and Robotics Teacher. Master's in Physics. State College of Campo Duque de Caxias, Antônio Olinto, Paraná, NRE of União da Vitória. Email: adriano.dubiel@escola.pr.gov.br

Paraná for his invaluable help and guidance throughout the development of this project.

5 Conclusion

The code concludes by printing a message indicating the number of robotics manuals downloaded and saved to Google Drive.

```
print('The program has downloaded', len(unique_urls), 'robotics manuals from SEED/PR for the year 2023 and saved them to Google Drive.')
```

6 Final Remarks

The presented code is an effective tool for collecting and organizing educational material in an automated manner, facilitating access to robotics manuals from SEED-PR. Adriano Dubiel, the author of this code, demonstrates a practical approach to automating tasks related to downloading and organizing educational resources.

7 References

– Adriano Dubiel - GitHub: https://github.com/ProfessorAdrianoDubiel/Robotica-Parana-2023
– Marcelo Santos - Federal University of Paraná
– Overleaf: https://www.overleaf.com