# Exercício webscrapy

## Exemplo de uso de IA integrada ao windsurf

* Pergunta:

@webscrapy\_41b\_v2.py , @webscraping\_41b\_resultado.csv , @webscrapy\_41b.py, quero melhorar a comunicação. como fazer para deixar mais claro o que foi realizado nessa sequencia de arquivos? uso o readme? Crio um novo documento? Também não quero borocratizar demais, ser prolixo demais. o projeto tem que ser mantido simples

* Resposta do modelo Gemini 2.5 Pro:

Of course. Improving documentation is a great way to make your project understandable and maintainable. Given your goal to keep things simple and clear, I believe the best approach is to enhance the existing [README.md](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/README.md:0:0-0:0) file. Creating a separate document might be overkill for this project's scale.

Here's my plan to help you with this:

1. I'll review the contents of `webscrapy\_41b.py` and [webscrapy\_41b\_v2.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b\_v2.py:0:0-0:0) to understand the evolution of your script.

2. I'll propose a clear and concise update to the [README.md](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/README.md:0:0-0:0) that explains the purpose of each file (`webscrapy\_41b.py`, [webscrapy\_41b\_v2.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b\_v2.py:0:0-0:0), and `webscraping\_41b\_resultado.csv`).

3. Once you approve the suggestion, I'll apply the changes to the [README.md](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/README.md:0:0-0:0) file.

Let's start by creating a plan and then reading the files.

I have reviewed both [webscrapy\_41b.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b.py:0:0-0:0) and [webscrapy\_41b\_v2.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b\_v2.py:0:0-0:0).

\* [webscrapy\_41b.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b.py:0:0-0:0) is the initial version, with all the logic in a single function.

\* [webscrapy\_41b\_v2.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b\_v2.py:0:0-0:0) is a much-improved version, refactored into clear functions with better documentation and structure.

To clearly explain the purpose of each file without overcomplicating your [README.md](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/README.md:0:0-0:0), I suggest replacing the current entry with a more descriptive list. This will clarify the evolution from the first script to the second and explain the output file.

Here is the change I propose for your [README.md](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/README.md:0:0-0:0):

I'll replace this section:

```markdown

- `trabalhos\_realizados/` → \*\*meus trabalhos e anotações\*\*

- [webscrapy\_41b.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b.py:0:0-0:0) - Trabalho de web scraping para coleta de dados

```

with this:

```markdown

- `trabalhos\_realizados/` → \*\*meus trabalhos e anotações\*\*

- \*\*Exercício de Web Scraping (4.1b):\*\*

- [webscrapy\_41b.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b.py:0:0-0:0): Versão inicial do script. Realiza a busca no Google e extrai os resultados em um único fluxo.

- [webscrapy\_41b\_v2.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b\_v2.py:0:0-0:0): Versão refatorada e melhorada, com funções separadas e mais robustez. \*\*(Recomendado para estudo)\*\*

- `webscraping\_41b\_resultado.csv`: Arquivo com os dados coletados pelo script (título e URL).

```

This approach keeps the documentation simple and centralized in the [README.md](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/README.md:0:0-0:0), explains the purpose of each file, and guides anyone reading it to the most relevant version of your script.

What do you think of this suggestion? If you agree, I will apply this change.

I've reviewed both the initial and the refactored versions of your web scraping script. To improve clarity without adding unnecessary complexity, I'll update the [README.md](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/README.md:0:0-0:0) to explain the purpose of each file in the `trabalhos\_realizados` directory.

Here's the plan:

1. I'll update the [README.md](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/README.md:0:0-0:0) to differentiate between [webscrapy\_41b.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b.py:0:0-0:0) (the initial version) and [webscrapy\_41b\_v2.py](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/trabalhos\_realizados/webscrapy\_41b\_v2.py:0:0-0:0) (the improved version), and also describe the `webscraping\_41b\_resultado.csv` output file.

2. I'll then mark the task as complete.

I'll now apply the changes to the [README.md](cci:7://file:///c:/Users/marce/OneDrive%20-%20mtegovbr/0001\_NOVA\_PASTA/DESENVOLVIMENTO\_PESSOAL/001\_mba\_ciencia\_dados\_enap/016\_introducao\_ciencia\_dados/mba\_enap\_introducao\_ciencia\_dados/README.md:0:0-0:0) file.