Relatorio_Web_Scraping

February 26, 2024

Isso representa ó código fonte de um site O código usado foi o mesmo fornecido pelo GitHub na descrição do vídeo, para melhor acompanhamento.

- tags são marcadores que especificam a estrutura do site e como as infomrações devem ser exibidas.
- </'nome_tag'> representa o fechamento de uma tag.
- Dentro da tag 'head' temos: ** tag 'meta' diz a respeito sobre os meta-dados do site. ** tag 'link' é resposável por importar o estilo para o site ** tag 'title' é responsável por
- A tag 'body' é responsável por mostrar a parte vísivel da página, o que realmente será exibido no site: ** tags 'hn' (h1,h2,etc) são tags de cabeçalho, onde sua ordem de importância decai de 1 (max) até 6 (min). ** tags 'div' permite imortar classes de estilo para o site, que por exemplo a de um cartão em '

  ' ** tags 'p' representam parágrafos de texto. ** tags 'a' permite que visitemos outra página, referenciando a página e o botão para ativar a ação.

[1]:
```
# pip install beautifulsoup4
# pip install lxml
# Estes pacotes servem para extrair dados de páginas web.
# A principal diferença entre parsers HTML e XML é que em XML é mais apropriado␣
 ↪para permitir que aplicações
# troquem e aloquem informações em sua estrutura em um jeito universalment␣
 ↪compreendido.

from bs4 import BeautifulSoup

with open('home.html', 'r') as html_file:
    content = html_file.read()
    soup = BeautifulSoup(content, 'lxml')
    # Método find() acha a primeira instancia da tag fornecida, porem␣
 ↪find_all() acha todas.
    tags = soup.find('h5')
    courses_html_tags = soup.find_all('h5')

    # for course in courses_html_tags:
    #     print(course.text)
    # print(courses_html_tags[0].text)
```

```
    course_cards = soup.find_all('div', class_='card')
    for course in course_cards:
        course_name = course.h5.text
        course_price = course.a.text.split()[-1]
        print(f'Course name: {course_name} | Price: {course_price}')
```

```
Course name: Python for beginners | Price: 20$
Course name: Python Web Development | Price: 50$
Course name: Python Machine Learning | Price: 100$
```