**Relatório 08 - Web Scraping com Python p/ Ciência de Dados**

Guilherme Loan Schneider

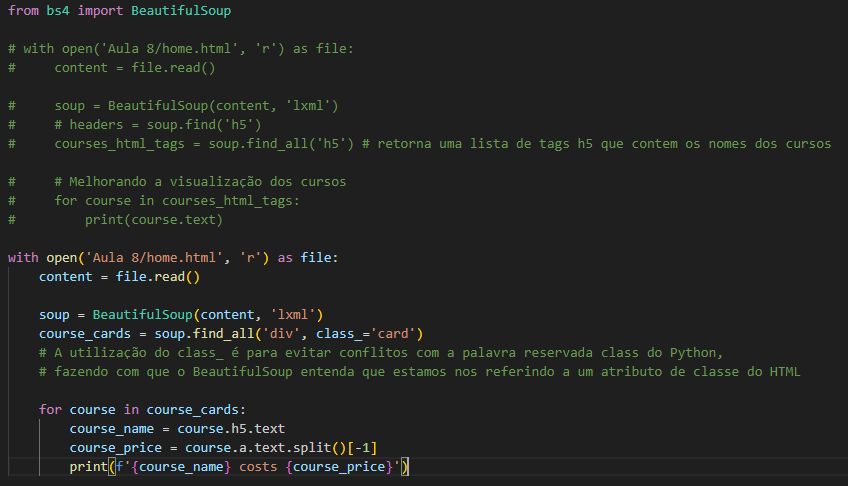
**Descrição da atividade**

A biblioteca BeautifulSoup é usada para analisar documentos HTML e XML, permitindo que você navegue pela estrutura da página de forma simples. Ela é útil para localizar e extrair elementos específicos de uma página, como títulos, links ou tabelas. Com o BeautifulSoup, é possível buscar elementos usando seletores como tags, classes e IDs, bem como percorrer a hierarquia de elementos HTML.

Por exemplo, se você tiver uma página HTML, o BeautifulSoup pode ajudar a encontrar todos os links (tags <a>) ou textos dentro de uma tag específica. Ele possui métodos como find() e find\_all() para buscas específicas, além de permitir manipulação da estrutura da página.

Outra importante biblioteca utilizada foi o “LXML”, que pode ser usado diretamente para manipulação de documentos ou como um parser para o BeautifulSoup. Além disso, o lxml suporta tanto HTML malformado quanto XML estruturado.

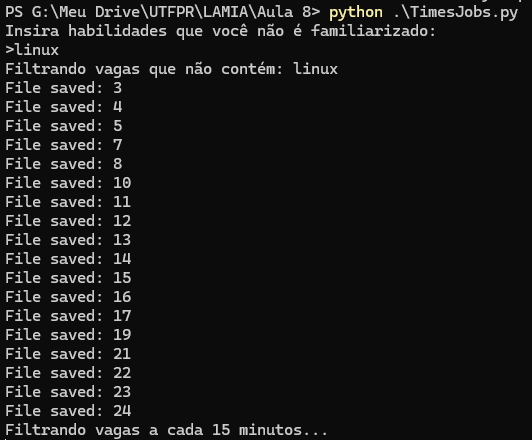
Na imagem abaixo, o autor utilizou a biblioteca BeautifulSoup em um arquivo HTML simples para demonstrar as formas de resgatar informações de uma página.



Na aplicação em um website real, foi fundamental utilizar a blbioteca “requests”, que obtêm o HTML de uma página para posteriormente analisá-lo com o BeautifulSoup. Ela simula como se fosse um indivíduo acessando um site e requisitando informações, como pesquisa, cliques em conteúdo do site, dentre outros.

Abaixo há o exemplo de uma aplicação em um site real chamado “TimesJobs”, utilizando a biblioteca requests, que acessa o site e resgata as informações definidas n código. Em seguida a execução do código que salva as aplicações procuradas em um arquivo TXT.





**Conclusões**

Foi possível compreender que a utilização dessas bibliotecas permite que sejam recuperadas informações de forma rápida de websites variados, basta apenas analisar de forma rápida como estão estruturadas as informações do site escolhido. Tudo isso combinado com a utilização das três bibliotecas citadas anteriormente a BeautifulSoup, LXML e Requests.

**Referencias**

[Web Scraping with Python - Beautiful Soup Crash Course](https://www.youtube.com/watch?v=XVv6mJpFOb0)