

 <p>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA PIAUI</p>	<p><b>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ</b></p> <p><b>Curso: ADS</b></p> <p><b>Disciplina: Programação para a Internet</b></p> <p><b>Professor: Ely</b></p>
--	--

Crie um pequeno mecanismo de busca na WEB utilizando a linguagem Python e as bibliotecas Requests e BeautifulSoup. O script deve funcionar através de uma chamada de função como:

*search(keyword, url, depth):*

- keyword é uma palavra-chave a ser buscada;
- url é um endereço inicial para a busca;
- depth é a profundidade da busca a partir da url.

Utilize sempre a biblioteca requests-cache ([https://requests-cache.readthedocs.io/en/latest/user\\_guide.html](https://requests-cache.readthedocs.io/en/latest/user_guide.html)) para armazenar páginas recentemente baixadas e melhorar o tempo de resposta de suas buscas.

Funcionamento da busca:

1. Para a URL, o buscador baixará a página usando a biblioteca requests:

```
from bs4 import BeautifulSoup
import requests
response = requests.get('http://www.ifpi.edu.br/')
```

Dentro da página, deve ser buscada a keyword digitada.

2. Caso encontre o texto, deve-se guardar um trecho do texto contendo a palavra. O trecho deve conter a palavra e, por exemplo, 20 caracteres antes e 20 depois da palavra.
3. Caso a profundidade da busca seja maior que zero, a página baixada deve ser passada para a biblioteca BeautifulSoup para retirar extrair os links e realizar uma busca recursiva nos links da página. Veja o exemplo abaixo em que o atributo href é extraído dos links de uma página:

```
html = BeautifulSoup(response.text, 'html5lib')
links = html.find_all('a')
for link in links:
    print(link['href'])
```

4. Implemente um mecanismo para Ranquear ou dar relevância nas páginas, exibindo primeiro as páginas principais e depois as secundárias de acordo com um critério a ser definido.
5. Pesquise como salvar as páginas em um banco pra evitar que requisições idênticas façam novas requisições, ainda que as páginas estejam em cache. Verificar última atualização da página ou data em que foi salva/pesquisada pela última vez pode ser uma forma de evitar buscas repetidas.