

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO Implementação de Linguagens de Programação – OPT011

Professor: Dr. Rogério Aparecido Gonçalves

ATIVIDADE: EXTRAÇÃO DE LINKS

Aluno: Eduardo Kurek RA: 2458586

Código fonte:

```
import re
def buscar_ocorrencias(arquivo, padrao):
    with open(arquivo, 'r', encoding='utf-8') as f:
        conteudo = f.read()
    regex = re.compile(padrao)
    ocorrencias = regex.findall(conteudo)
    if ocorrencias:
        print(f"Ocorrências encontradas ({len(ocorrencias)}):")
        for ocorrencia in ocorrencias:
            print(ocorrencia)
    else:
        print("Nenhuma ocorrência encontrada.")
tag = input("Digite a tag que deseja buscar: ")
arquivo_txt = 'utfpr-cm.html'
padrao_regex = rf'<{tag}\b.*?>.*?</{tag}>' # 0 asterisco não pode
buscar_ocorrencias(arquivo_txt, padrao_regex)
```

Resultados:

```
$ py codigo.py
Digite a tag que deseja buscar: a
Ocorrências encontradas (126):
a class="skiplink" href="https://www.utfpr.edu.br/campus/campomourao#view">Ir para o conteé<
do</a>
a class="skiplink" href="https://www.utfpr.edu.br/campus/campomourao#navigation">Ir para a
navegação</a>
a class="skiplink" href="https://www.utfpr.edu.br/campus/campomourao#footer">Ir para o roda
pé</a>
<a accesskey="1" href="https://www.utfpr.edu.br/campus/campomourao#conteudo">Ir para o conte
údo</a>
<a href="https://gov.br/" class="pic-gov" title="GovBR">GovBR</a>
a href="http://www.gov.br/secom/pt-br/acesso-a-informacao/comunicabr" class="link-barra lin"
k-externo-barra">Comunica BR</a>
<a href="http://www.gov.br/acessoainformacao/" class="link-barra link-externo-barra">Acesso
à informação</a>
a href="https://www.gov.br/pt-br/participacao-social/" class="link-barra link-externo-barra"
">Participe</a>
a href="http://www4.planalto.gov.br/legislacao/" class="link-barra link-externo-barra">Legi>
slação</a>
<a href="http://www.gov.br/pt-br/orgaos-do-governo" class="link-barra link-externo-barra">Ór
gãos do Governo</a>
a ria-label="login" href="https://www.utfpr.edu.br/campus/campomourao/login-menu">Entrar</
a>
<a title="Início" href="https://www.utfpr.edu.br/" class="w-fit hidden sm:block"><svg width=
"424" height="92" viewBox="0 0 424 92" shape-rendering="auto" fill="none" xmlns="http://www.
```

w3.org/2000/svg"><path class="fill-[#0D1117] dark:fill-[#F8FAFA]" d="M142.941 75.1993H144.82