

## PyGrep

Uma das funcionalidades que usamos bastante durante a programação é procurar sequências de texto em arquivos dispersos pelo nosso computador, em diferentes arquivos. Um dos mais clássicos programas que realiza este tipo de tarefa é o `grep`, que está presente na maioria dos sistemas operacionais `unix/linux`. Este software faz uso das técnicas de expressões regulares para encontrar padrões textuais dentro do arquivo mencionado, versões mais avançadas podem pesquisar por conteúdo textual em uma pasta repleta de arquivos em velocidade bastante rápida.

Nesta aula, vamos implementar uma versão simplificada deste programa para pesquisar em arquivos e pastas, trechos textuais que contenham uma string específica informada pelo usuário no momento da invocação do script.

Vamos começar, o primeiro passo é criar uma estrutura de pastas que vai servir como o nosso modelo de testes, crie os seguintes arquivos e pastas conforme a estrutura abaixo, em uma pasta a sua escolha.

```
/files
  /docs
    doc.txt
  file.txt
pygrep.py
```

Em seguida, abra o arquivo `pygrep.py`, e vamos começar a estruturar o nosso script. O primeiro passo é obter os valores que serão informados pelo usuário como argumentos do programa, para isso, precisamos importar o módulo `sys`, e acessar a propriedade `argv`, que contém a lista de argumentos informados pelo executor do script, na ordem de inserção, veja o seguinte comando:

```
python pygrep.py data arquivo files/file.txt
```

Para este comando, a lista `sys.argv` terá como itens os seguintes valores:

0. `pygrep.py`;
1. `data`;
2. `arquivo`;
3. `files/file.txt`;

Com esse conhecimento em mente, podemos escrever o nosso script com uma consideração importante: serão necessários no mínimo três argumentos para que o programa funcione corretamente, caso estes três argumentos não sejam informados vamos dar uma mensagem de erro para o usuário.

```
import sys

argv_len = len(sys.argv)

if argv_len > 3:
    search_string = sys.argv[1]
    search_type = sys.argv[2]
    path = sys.argv[3]
else:
    sys.exit("Argumentos inválidos. Utilize string_de_busca tipo_de_busca(arquivo/pasta) caminho.")
```

Agora com as variáveis importantes para o nosso trabalho, podemos começar a estruturar o funcionamento do programa, o primeiro passo é verificar o tipo de busca, se é uma busca em um arquivo específico ou uma pasta específica. Caso não seja especificado conforme o esperado, vamos dar uma mensagem de erro.

```
if search_type == 'arquivo':
    search_file(path)
elif search_type == 'pasta':
    search_folder(path)
else:
    sys.exit("Tipo de busca não reconhecido pelo sistema, utilize arquivo ou pasta.")
```

Ok, agora precisamos criar as duas funções que estamos utilizando no trecho acima, a primeira, chamada `search_file(path)`, vai buscar conteúdo dentro de um arquivo específico. A função `search_folder(path)`, vai buscar por todos os arquivos dentro daquela pasta e subpastas. Vamos definir cada uma delas abaixo, lembrando de incluir os imports específicos.

```
import io
...
def search_file(file_path):
    with io.open(file_path) as file:
        if search_string in file.read():
            print("Texto encontrado no arquivo localizado em:", file_path)
import os
...
def search_folder(folder_path):
    with os.scandir(folder_path) as entries:
        for entry in entries:
            if entry.is_file():
                search_file(entry.path)
            if entry.is_dir():
                search_folder(entry.path)
```

Com isso temos agora as definições das funções e finalmente completamos todo o script, podemos testar informando um trecho do texto presente nos arquivos criados no começo deste projeto.

O script final vai ficar mais ou menos assim:

```
import sys
import io
import os

argv_len = len(sys.argv)

if argv_len > 3:
    search_string = sys.argv[1]
    search_type = sys.argv[2]
    path = sys.argv[3]
else:
    sys.exit("Argumentos inválidos. Utilize string_de_busca tipo_de_busca(arquivo/pasta) caminho.")

def search_file(file_path):
    with io.open(file_path) as file:
        if search_string in file.read():
            print("Texto encontrado no arquivo localizado em:", file_path)

def search_folder(folder_path):
    with os.scandir(folder_path) as entries:
        for entry in entries:
            if entry.is_file():
                search_file(entry.path)
            if entry.is_dir():
```

```
        search_folder(entry.path)

if search_type == 'arquivo':
    search_file(path)
elif search_type == 'pasta':
    search_folder(path)
else:
    sys.exit("Tipo de busca não reconhecido pelo sistema, utilize arquivo ou pasta.")
```