

Projeto – Busca por *strings* em documentos

Definições iniciais

A implementação do projeto poderá ser realizada individualmente ou em equipes formadas por até três integrantes. Cabe ressaltar que o projeto final também poderá ser realizado com esta mesma configuração, então as equipes poderão manter suas formações para implementação da continuidade da aplicação, caso todos os integrantes estejam de acordo.

Este projeto possui **peso de 30% na avaliação do curso**. Importante lembrar que a emissão do certificado e a **atribuição das horas complementares está condicionada à obtenção da média 6**.

O prazo para entrega deste projeto é dia 01/09/2021, às 17h. As instruções para entrega são apresentadas no final deste enunciado.

Descrição da Aplicação

O projeto consiste na implementação de uma aplicação para localizar palavras informadas pelo usuário em documento PDF. A aplicação deverá solicitar que o usuário informe uma ou mais palavras em campo ou entrada específico. A entrada deverá seguir as seguintes regras:

- As palavras que deverão ser localizadas no texto serão informadas em letras minúsculas.
- No caso de múltiplas palavras deverá ser informado o operador lógico entre elas. Os operadores lógicos aceitos serão AND e OR.
- Os operadores deverão ser escritos em letras maiúsculas, enquanto as palavras consultadas em letras minúsculas.

*** Um exemplo de entrada é: desenvolvimento AND aplicação.**

A aplicação deverá retornar os seguintes resultados, com base na consulta realizada.

- Caso o operador OR seja utilizado, a aplicação deverá retornar o número de ocorrências de cada uma das palavras consultadas.
- Caso o operador AND seja utilizado, a aplicação deverá retornar se as duas palavras foram encontradas no texto, juntamente com o número de ocorrências de cada uma delas.

É importante frisar que as palavras poderão constar no documento em mais de um formato. Por exemplo, a palavra **aplicação** pode constar também como **Aplicação**, **APLICAÇÃO**, **aplicacao**, etc.

Na avaliação desta etapa do projeto não serão utilizadas entradas contendo operadores AND e OR em conjunto, ou seja, a entrada usará apenas um dos operadores. No entanto, a entrada poderá conter múltiplas palavras e esse cenário deve ser tratado pelos desenvolvedores.

EXTRA: Caso alguma equipe implemente o tratamento do uso conjunto dos operadores, juntamente com parênteses, deverá avisar os professores para a realização de testes específicos. Esta implementação será considerada um diferencial do projeto.

Além da entrada para que o usuário informe a string que deverá ser pesquisada no documento, a aplicação deverá solicitar que o usuário informe o documento que será utilizado na pesquisa. A aplicação poderá usar uma pasta default, na qual o usuário deverá copiar os documentos que deseja utilizar nas consultas, informando o nome do documento para a busca atual. Outra alternativa é oferecer ao usuário a possibilidade de inclusão do caminho no nome do documento. Caso a equipe opte pelo uso de interface gráfica, existem componentes associados à seleção de arquivos.

Por fim, a aplicação deverá armazenar em um arquivo texto o histórico das pesquisas realizadas. O arquivo deverá conter o número da consulta, considerando todas já realizadas, o nome do documento pesquisado, a string de busca e o número de ocorrências associado com cada palavra. Abaixo é apresentado um exemplo de como o histórico pode ser mantido:

Número da consulta: 10

Nome do documento: Enunciado.pdf

String de busca: desenvolvimento AND aplicação

Ocorrências: desenvolvimento (1), aplicação (3)

Entrega da aplicação

A entrega do projeto será realizada através do GitHub. Para isso, na segunda aula, os professores disponibilizarão um Form para que a equipe preencha seus integrantes e o endereço do GitHub. A equipe poderá utilizar a conta do GitHub de um dos seus integrantes.

Apesar da equipe incluir seus integrantes do Form, solicitamos que os créditos também sejam incluídos na aplicação desenvolvida.

No dia 01/09/2021, a partir das 17h, os professores acessarão os endereços do GitHub informados e farão um clone das aplicações para avaliação.

Método de avaliação

Os critérios de avaliação do projeto são apresentados a seguir. Também informamos o peso que será considerado para cada item na correção:

1. Contagem correta das palavras que compõem a string de busca (peso: 5 pontos);
2. Tratamento correto do operador AND (apresentar se todas as palavras estão presente no documento) (peso: 1 ponto);
3. Tratamento da variação das palavras em relação aos acentos e cedilha (peso: 2 pontos);
4. Gravação de todas as informações associadas ao cache de consultas (peso: 2 pontos);
5. Implementação dos critérios EXTRA para tratamento de múltiplos operadores e ordem de precedência (parênteses) (valor extra: 1 ponto na média final, considerando que a nota máxima do curso é 10).