

Bacharelado em Ciência da Computação

FACOM31303 Algoritmos e Estruturas de Dados 2

Projeto Final

Profa. Maria Camila Nardini Barioni

APLICAÇÃO DE PESQUISA

1. Informações Básicas

Este trabalho tem por objetivo reforçar o conhecimento dos discentes com relação aos algoritmos para modelagem de problemas de busca/pesquisa.

- Data da entrega: 12/04/2025 na atividade criada no Teams
- Grupo de até 3 (três) alunos(as)
- Linguagem de programação a ser usada: C
- O que deve ser entregue:
 - O trabalho deve ser entregue na forma de uma apresentação em PDF que contenha a explicação das soluções adotadas pelo grupo para cada item da descrição do problema com ilustrações, uma visão geral do código com a indicação das referências usadas e casos de uso (exemplos de execução). **No primeiro slide da apresentação devem constar os nomes e números de matrícula dos discentes e o link para o repositório em que o código está armazenado (Replit.com).** A apresentação deverá estar em formato PDF e deve ser preparada levando em consideração que cada grupo terá 15 minutos para apresentar e demonstrar o trabalho.
 - O código deverá estar disponível em um repositório online e seu endereço indicado no documento da apresentação. É responsabilidade do grupo certificar-se que o código estará visível para a professora e monitor para a avaliação e execução;
 - Para o dia da apresentação, cada grupo deverá preparar um roteiro para demonstrar cada uma das funcionalidades solicitadas no detalhamento.

2. Detalhamento

Objetivo: Projetar e implementar um repositório de citações de filmes..

Descrição do Problema:

Cada entrada no repositório de citações de filmes é uma tripla: **palavra**, dados das **citações** em que essa palavra ocorre (lista com os offsets dos inícios das linhas que contém os dados das citações em que a palavra ocorre), e a contagem da **frequência** da palavra no repositório.

Exemplo, considerando a palavra *life*:

Life
1821, 2041, 5614, ...
12

Bacharelado em Ciência da Computação

FACOM31303 Algoritmos e Estruturas de Dados 2

Projeto Final

Profa. Maria Camila Nardini Barioni

Faça um programa que leia arquivos texto de citações de filmes e organize, em ordem alfabética, as palavras (com mais de 3 caracteres) contidas nas citações e as suas frequências em estruturas de dados. Espaços em branco e sinais de pontuação devem ser descartados, e as palavras devem ser mantidas todas em minúsculo.

O processamento das palavras das citações de filmes (e suas informações adicionais) deverá realizar-se de três formas distintas:

1. Pesquisa Binária (usando vetor).
2. Árvore Binária de Busca sem balanceamento.
3. Árvore AVL.

O programa deve permitir que:

A. O usuário escolha um arquivo de texto para ser carregado nas estruturas de dados, considerando cada um dos 3 casos citados. Após a carga dos dados nas estruturas, deve ser exibido o tempo para cada uma. A função de inserção deverá ser modificada para considerar palavras já processadas da seguinte forma, se a palavra sendo processada já estiver inserida na estrutura de dados, deve-se atualizar a contagem geral de frequência da palavra no repositório de citações de filmes.

B. O usuário escolha uma palavra para visualizar as informações das citações que contém essa palavra (para cada citação exibir: citação, filme, ano) e a frequência da palavra no repositório, caso a palavra tenha sido processada. O resultado da busca deve ser exibido para as três situações programadas, assim como o tempo de processamento da consulta.

C. O usuário faça busca por frequência. Neste caso, deve ser construída uma árvore AVL considerando-se a organização dos dados por frequência das palavras no repositório para que a busca seja eficiente.

Observações:

- As funções devem estar organizadas em arquivos modularizados (.h e .c)
- O código deverá estar corretamente indentado.
- Utilize o arquivo “movie_quotes.csv” na realização do trabalho. Esse arquivo está disponível na equipe da disciplina no MS Teams e deve ser incluído no seu repositório do código no Replit. Para a apresentação final o grupo deverá executar o código considerando o arquivo completo. Para a correção do código no Replit, o grupo pode incluir também um arquivo com um subconjunto dos dados do arquivo original para permitir a execução do mesmo.

Bacharelado em Ciência da Computação

FACOM31303 Algoritmos e Estruturas de Dados 2

Projeto Final

Profa. Maria Camila Nardini Barioni

- Sugiro explorar as funções `fgets()` e `ftell()` para a implementação da lista de offsets indicada na descrição do problema.

Critérios de avaliação:

- (a) Apresentação: objetividade, clareza, organização do texto/algoritmo/código, formatação/ilustração, roteiro de demonstração.
- (b) Programa: qualidade dos dados informados / qualidade da impressão visão do percurso / modularização / uso correto das estruturas de dados / legibilidade e consistência na codificação das funções de manipulação do grafo, da árvore e do algoritmo de ordenação/ documentação do código.

ⁱ Arquivo adaptado de <https://www.kaggle.com/datasets/mexwell/movie-quotes/data>