

Bacharelado em Ciência da Computação

GBC034 Algoritmos e Estruturas de Dados 2

Trabalho 2

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: 05/11/2023
- Grupo de até 2 (dois) alunos(as)
- Linguagem de programação a ser usada: C
- O que deve ser entregue: Link para o repositório do código (Replit.com) e relatório em formato .pdf contendo:
 - Descrição do problema e da solução/abordagem usada pela dupla para tratar o problema
 - Descrição das estruturas utilizadas
 - Código das principais funções
 - Exemplos rodados
 - Endereço do repositório do código
 - Referência bibliográfica usada

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

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.

Bacharelado em Ciência da Computação

GBC034 Algoritmos e Estruturas de Dados 2

Trabalho 2

Profa. Maria Camila Nardini Barioni

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

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