

Trabalho Prático II (10 Pontos) – Até 3 alunos por grupo

Como trabalho prático II da disciplina de AED3 cada grupo terá duas opções de atividade.

Opção 1) Sistema de Arquivo Invertido e Pesquisa

Esta opção consiste em construir um sistema de arquivo invertido para indexação e pesquisa de palavras em arquivos texto. Essencialmente, o sistema terá as seguintes opções:

1) Indexar Arquivos Textos: o usuário deverá informar um ou mais arquivos texto a serem indexados. Após informar o arquivo, o sistema deverá proceder com a geração/atualização de um arquivo de índices invertidos. Cada linha do arquivo terá um conjunto de tuplas no seguinte formato:

$$\text{Palavra } \langle A_1, X_1, Y_1 \rangle \langle A_2, X_2, Y_2 \rangle \dots \langle A_Z, X_N, Y_M \rangle$$

Onde A_k corresponde ao nome do arquivo onde a palavra acontece, X_i e Y_i correspondem a linha e coluna onde uma determinada palavra aconteceu no arquivo A_k .

O sistema deverá estar preparado para que o usuário possa adicionar novos arquivos textos a qualquer momento, visando atualizar o arquivo de índices com novos conteúdos.

2) Pesquisa por palavras: o usuário poderá realizar uma pesquisa por palavras considerando as seguintes operações:

- Se o usuário informar mais de uma palavra entre aspas (ex. "Sistemas Inteligentes") a busca deverá retornar os trechos do texto onde ocorreram as duas palavras em sequência ("Sistemas" e "Inteligentes").
- Se as palavras não estiverem entre aspas (ex. Sistemas Inteligentes), o sistema deverá retornar trecho onde acontece a primeira palavra ("Sistemas") dos arquivos que contenham as duas palavras.

Entende-se por trecho, o conjunto de 10 palavras anteriores e 10 palavras posteriores ao termo procurado, dispostos no arquivo original.

Se as palavras não forem encontradas no arquivo indexado, uma mensagem deverá ser mostrada ao usuário indicando que não é possível

3) Definir arquivo de *stopwords*. O usuário poderá informar um arquivo texto formado por palavras que não devem ser indexadas (ex. caracteres de pontuação, palavras proibidas, preposições indesejáveis, etc).

Se o aluno achar conveniente, estruturas de dados em memória primária poderão ser utilizadas para auxiliar o processo de indexação, como por exemplo, listas encadeadas.

Opção 2) Sistema de Cadastro de Filmes - Continuação

No trabalho prático 1 você desenvolveu um sistema com opções de inserção, leitura/busca, atualização e exclusão, também conhecido como CRUD (Create Read, Update, Delete).

Neste trabalho você deverá implementar um relacionamento 1:N. Isso exigirá a criação do CRUD (Create Read, Update, Delete) de uma segunda entidade. Nesta tarefa, vocês devem implementar um CRUD de gêneros de filmes (Ação, Comédia, Guerra, Policial, Terror, etc.). Essa entidade precisará apenas de um atributo que contenha o nome do gênero.

Em seguida, vocês devem alterar a entidade de filmes para que contenha mais um atributo que contenha o ID do gênero a que cada filme pertence. A consistência deve ser mantida em todas as operações de inclusão, alteração e exclusão de filmes.

Também deve ser possível emitir a lista de filmes de uma determinada categoria.

Roteiro para Implementação (sugestão)

- 1) Atualizem o CRUD de filmes (TP1) para que usem índices.
- 2) Acrescentem, à entidade Filme, um atributo idGenero.
- 3) Criem um CRUD para os gêneros. É um CRUD simples e será bem semelhante ao CRUD anterior.
- 4) Atualizem o menu principal para que seja possível acessar todas as operações dos CRUDs de filmes e gêneros.

Neste momento, vocês têm dois CRUDs que atendem às operações necessárias, mas ainda não é possível garantir a consistência entre eles, isto é, que um filme só seja cadastrado em gêneros válidos. Para isso, precisaremos, a cada vez que um idGenero for digitado, ver se é um ID válido. Da mesma forma, na hora de apresentar um filme, precisaremos apresentar também o nome do gênero a que ele pertence.

5) Nas funções de inclusão e de alteração de filmes, após o usuário digitar o ID do gênero, usem a função de busca do arquivo de gêneros e chequem se algum gênero válido é retornado. Se houver um gênero válido, apresentem o nome desse gênero. Caso contrário, informem que o gênero não existe e solicitem a digitação de um novo ID de gênero.

6) Nas funções de busca e de listagem de filmes será também necessário usar dados do gênero. Assim, a apresentação dos dados do filme deverá ser feita localmente, atributo por

atributo. Nesse processo, ao escrever o `idGenero`, façam a mesma busca do passo anterior e informem o nome do gênero do filme.

Ainda precisamos garantir que um gênero não será excluído se o seu **ID** estiver sendo usado por algum filme. Apesar de ainda não termos estudado uma estrutura de dados eficiente para isso, vamos fazer o processo da forma mais simples (e lenta).

7) Criem uma função que receba um **ID** de gênero como parâmetro e que use esse **ID** para filtrar os filmes listados, isto é, esse novo método retornará uma lista dos filmes que pertencem a um determinado gênero. Assim, o gênero só poderá ser excluído se a lista retornada estiver vazia.