

Trabalho 2

Disciplina: Estrutura de Dados II

Prof. Dr. Lucas C. Ribas

1 Organização de Arquivos e Indexação em Árvore B^+

Um colecionador deseja manter um cadastro de seus filmes em um arquivo. Para cada filme, o colecionador deseja manter os seguintes atributos:

- Código do filme (composição das três primeiras letras do último sobrenome do diretor e dois últimos dígitos do ano de lançamento do filme, e.g., WEL41) — esse campo é chave, não existirão valores idênticos.
- Título em português.
- Título original (se o título em português for o mesmo, deve-se utilizar a palavra Idem).
- Diretor (sobrenome e nome do diretor, e.g.: Welles, Orson).
- Ano de lançamento (e.g., 1941).
- País (país no qual o filme foi produzido).
- Nota (inteiro entre 0 e 9 com a nota dada pelo colecionador ao filme).

Tarefa

Desenvolva um programa que permita ao colecionador controlar seus filmes. O programa deverá permitir:

1. Inserir um novo filme no catálogo.
2. Remover um filme a partir da chave primária.
3. Modificar o campo nota de um filme a partir da chave primária.
4. Buscar filmes a partir da (a) chave primária (Árvore B^+) ou do (b) título em português (índice secundário).
5. Listar todos os filmes no catálogo ordenados pela chave primária usando o conjunto de sequências (lista de folhas) da Árvore B^+ .

Para realizar essa tarefa será necessário organizar 3 arquivos distintos:

1. um arquivo de dados que conterá todos os registros,
2. um arquivo de índice primário que será uma **Árvore** B^+ ,
3. um arquivo de índice secundário para o título em português.

Estrutura do arquivo de dados

O arquivo de dados deve ser ASCII (arquivo texto) e organizado em registros de tamanho fixo de 192 bytes (caracteres). Os campos de títulos, diretor e país devem ser de tamanho variável. Os demais campos devem ser de tamanho fixo: ano (4), nota (1) e chave primária (5). A soma de bytes dos campos fornecidos (incluindo os delimitadores necessários) nunca irá ultrapassar 192 bytes. Os campos do registro devem ser separados pelo caractere delimitador @ (arroba). Cada registro terá 7 delimitadores, mais 10 espaços ocupados pelos campos de tamanho fixo. Você poderá fixar um tamanho máximo para cada campo: título, título original, diretor e país de forma a garantir que ocupem, juntos, um máximo de 175 bytes (192-7-10). Caso o registro tenha menos de 192 bytes, o espaço adicional deve ser marcado de alguma forma a completar os 192 bytes. Segue abaixo um exemplo com três registros. Os espaços adicionais foram preenchidos com o caractere #.

```
ROC64@Deus e o Diabo na Terra do Sol@Idem@Rocha, Glauber@1964@Brasil@8@#####  
#####  
WEL41@Cidadão Kane@Citizen Kane@Welles, Orson@1941@Estados Unidos@9@#####  
#####  
KUB64@Dr. Fantástico@Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb@Kubrick, Stanley@1964@Estados Unidos@9@#####
```

Figura 1: Exemplo de arquivo de dados com registros e campos separados pelo caractere @.

Note que não há quebras de linhas no arquivo (elas foram inseridas aqui apenas para exemplificar a sequência de registros).

O arquivo de dados não deverá conter cabeçalho e deverá se chamar `movies.dat`. Instruções para as operações com registros:

- **Inserção:** cada filme inserido no catálogo deve ser inserido no final do arquivo de dados e atualizado nos índices. Chave também deve ser inserida no índice primário (**Árvore** B^+).
- **Remoção:** o registro deverá ser localizado pela chave primária acessando o índice primário (**Árvore** B^+). A remoção deve colocar os caracteres *| nas duas primeiras posições do registro removido. O espaço do registro removido **não** deverá ser

reutilizado para novas inserções. Chave também deve ser removida no índice primário (**Árvore B^+**).

- **Atualização:** o registro deverá ser localizado acessando o índice primário (**Árvore B^+**). A nota deverá ser atualizada no registro na mesma posição em que está (não deve ser feita remoção seguida de inserção).

Índices

Dois arquivos com índices serão criados:

- `ibtree.idx`: índice primário, contendo a chave primária e o RRN do respectivo registro. Este índice deve ser organizado em uma **Árvore B^+** que ficará armazenada no arquivo. Isto é, as páginas devem ser carregadas em memória principal somente quando necessário.
- `ititle.idx`: índice secundário, contendo o título e a chave primária do respectivo registro, ordenado pelo título.

A implementação do índice e **Árvore- B^+** é livre. No entanto, é importante lembrar que a **Árvore- B^+** deve estar sempre em arquivo.

Interação com o usuário

O programa deve permitir interação com o usuário pelo console/terminal (modo texto). As seguintes operações devem estar disponíveis:

- **Inserção de filme**: O usuário deve ter a capacidade de adicionar um novo filme ao sistema. O processo de inserção segue os seguintes passos:
 1. **Coleta de Informações:** O programa deve solicitar ao usuário as seguintes informações sobre o filme:
 - **Título em Português**
 - **Título Original**
 - **Diretor**
 - **Ano de Lançamento**
 - **País**
 - **Nota**
 2. **Geração da Chave:** A chave não é fornecida pelo usuário. O programa deve gerar automaticamente uma chave única para o novo registro de filme.
 3. **Verificação da Chave na Árvore B^+ :** Antes de inserir o filme no arquivo de dados, o programa deve verificar, através de uma busca na **Árvore B^+** , se a chave primária já existe.
 - **Se a chave existir:** O programa deve mostrar uma mensagem de erro e não prosseguir com a inserção.

- **Se a chave não existir:** Prossiga para o próximo passo.
- 4. **Inserção de Dados:** Caso a chave não exista, o programa deve:
 - Inserir o registro do filme no arquivo de dados.
 - Atualizar o índice secundário.
 - Inserir a chave na Árvore B^+ .
- **Correção da Nota de Filme:** O usuário deve conseguir corrigir a nota de um filme já existente no sistema. O processo de correção segue os seguintes passos:
 1. **Identificação do Filme:** O programa deve solicitar ao usuário a chave do filme cuja nota ele deseja corrigir.
 2. **Busca na Árvore B^+ :** Utilizando a chave, o programa realiza uma busca na Árvore B^+ para localizar o filme.
 - **Se o filme não for encontrado:** O programa deve mostrar uma mensagem indicando que o filme não existe no sistema e não prosseguir com a correção.
 - **Se o filme for encontrado:** Prossiga para o próximo passo.
 3. **Solicitação da Nova Nota:** O programa deve agora solicitar ao usuário que informe a nova nota para o filme.
 4. **Atualização de Dados:** Após receber a nova nota:
 - O programa atualiza a nota do filme no arquivo de dados.
 5. **Confirmação ao Usuário:** Uma vez que a nota foi atualizada com sucesso, o programa deve informar ao usuário que a nota do filme foi corrigida.
- **Remoção:** O usuário deve conseguir remover um filme do sistema. O processo de remoção segue os seguintes passos:
 1. **Coleta da Chave:** Para remover um filme, o programa deve solicitar ao usuário o campo chave correspondente ao filme desejado.
 2. **Verificação da Existência do Filme em Árvore B^+ :**
 - **Se o filme não existir:** O programa deve informar ao usuário que o filme solicitado não foi encontrado no sistema.
 - **Se o filme existir:** Prossiga para o próximo passo.
 3. **Remoção do Filme:**
 - Marque o registro correspondente como removido no arquivo de dados.
 - Remova a entrada associada ao filme no índice primário (Árvore B^+).
 4. **Confirmação:** Após a remoção bem-sucedida, o programa pode opcionalmente informar ao usuário que o filme foi removido com sucesso.
- **Busca de filme:** O usuário deve ser capaz de buscar por um filme.
 - **Pela chave:** O programa deve solicitar ao usuário a chave do filme desejado.

- * Se o filme não for encontrado utilizando a Árvore B^+ (índice primário), o programa deve informar que o filme não existe.
 - * Se o filme for encontrado, todos os seus dados devem ser impressos na tela em um formato legível e organizado.
- **Pelo título:** O programa deve solicitar ao usuário um título.
 - * Se não houver filmes com o título informado, o programa deve informar ao usuário.
- **Listagem de filmes**: O usuário tem a opção de listar filmes de acordo com os critérios abaixo:
 - **Por chave maior que um valor especificado (range):** O programa deve solicitar ao usuário um valor de chave.
 - * O programa deve listar todos os filmes cujas chaves sejam maiores do que o valor fornecido pelo usuário.
 - * Esta listagem deve utilizar a Árvore B^+ (índice primário) para identificar e recuperar os filmes que atendam a este critério. Buscar pela chave na árvore e depois usar o conjunto de sequências para pegar as demais.
 - * Se nenhum filme atender ao critério, o programa deve informar ao usuário que não existem filmes com chaves maiores que o valor fornecido.
 - **Listar todos os filmes:**
 - * O programa deve exibir todos os filmes disponíveis no banco de dados.
 - * Cada registro de filme deve ser impresso na tela em ordem (chave), de maneira formatada e organizada. Usar o conjunto de sequência da Árvore B^+ .
- **Finalizar a execução**: O usuário deve ser capaz de encerrar a execução do programa. Ao final da execução, atualize os índices no disco, feche todos os arquivos e libere toda a memória alocada pelo seu programa.

Implementação

O seu programa deve ter as seguintes funcionalidades implementadas, que podem, opcionalmente, serem colocadas em uma biblioteca de manipulação de arquivos:

- Uma estrutura de dados para armazenar os índices na memória principal.
- Verificar se o arquivo de dados existe.
- Verificar se o índice primário (Árvore B^+) existe.
- Verificar se o índice secundário existe.
- Criar o índice secundário: deve refazer o índice secundário a partir do arquivo de dados e substituir, caso haja, um índice existente no disco.
- Carregar página/nó do disco para a memória principal.

- Salvar página/nó da memória principal para o disco.
- Inserir um registro: modificando o arquivo de dados no disco, e os índices.
- Alterar um registro: modificando o arquivo de dados no disco, e os índices.
- Remover um registro: modificando o arquivo de dados no disco, e os índices.
- Buscar pela chave primária.
- Buscar pela chave secundária.

2 Avaliação

Para avaliação deste trabalho será utilizado a tabela de pontuação abaixo. Considere que cada aluno/dupla inicia com um total de pontos $P = 100$, pontos serão retirados/adicionados conforme os critérios listados:

Nº	Critério	Pontos
1	Trabalho individual e implementou/funcionou remoção na Árvore B^+	+10
2	Trabalho em dupla e não implementou/funcionou remoção na Árvore B^+	-10
3	Código não executa	-50
4	Atraso na entrega	-10 por dia
5	Código mal organizado e documentado	-10
6	Resultado/operação incorreta ou inesperada durante a execução	-10 por erro
7	Resposta imprecisa na entrevista	-10 por resposta
8	Não implementação/funcionamento de operação	-15 por operação
9	Não implementação/utilização de Árvore B^+	-100
10	Não realizar a apresentação	-100
11	Não entregar/realizar o trabalho	-100
12	Cópia de código fonte de qualquer origem	-100

*Ajustes podem ser feitos na tabela. Não necessariamente todos critérios são cumulativos, eles serão aplicados de acordo com cada caso. Por exemplo, o não funcionamento da busca por chave significa que essa opção não está disponível para o usuário, aplicando-se o critério 8. Por outro lado, se uma busca que esteja implementada (de modo aparentemente correta) retornar um resultado errado, aplica-se o critério 6. É importante notar que código que não executa, por estar totalmente incompleto por exemplo, pode se enquadrar também no critério 13.

- A pontuação inicial para todos é $P = 100$.
- A pontuação não será inferior a $P = 0$.
- Após a correção, apresentação e aplicação dos critérios da tabela, a nota do segundo trabalho será $T_2 = P/10$.

- A nota final dos trabalhos no semestre será: $T = \frac{(T_1*1)+(T_2*2)}{3}$. O valor de T tem peso de 30% na nota final do semestre.
- Códigos do Trabalho 1 podem ser reaproveitados neste trabalho.
- O projeto deverá ser entregue apenas pelo classroom até a data limite de **28/11**.
- Submissão deve ser realizado no seguinte formato: arquivos de dados e códigos **em C** compactado como NomeSobrenome_RA.zip. Cada integrante da dupla (quando for o caso) deve entregar individualmente.
- Dúvidas deverão ser esclarecidas pelo classroom. Preferencialmente, envie a dúvida para que todos possam ver, pois pode auxiliar outros colegas também. Uma aula de dúvidas também será agendado em Novembro.
- Entrevistas serão realizadas para correção do trabalho no dia **29/11**. Entregar o trabalho não garante nota. É preciso demonstrar que você fez e sabe o que fez.