

**Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Departamento de Ciências de Computação
Disciplina de Organização de Arquivos (SCC0215)**

Docentes

Prof. MSc. Anderson Canale Garcia

andersongarcia@usp.br

Profa. Dra. Cristina Dutra de Aguiar

cdac@icmc.usp.br

Monitores

Beatriz Aimee Teixeira Furtado Braga

beatriztfb@usp.br ou telegram: @bia_aimee

Lucas Piovani

lucaspiovani@usp.br ou telegram: @lucaspiovani

Vitor Augusto Paiva de Brito

vitor_brito@usp.br ou telegram: @vitorpbrito

Trabalho Prático 1

Este trabalho tem como objetivo inserir e remover dados de um arquivo binário, bem como desenvolver funcionalidades para recuperar dados desse arquivo usando um índice.

O trabalho deve ser feito por, no máximo, 2 alunos da mesma turma. Os alunos devem ser os mesmos do primeiro trabalho prático. Quaisquer mudanças devem ser devidamente informadas. A solução deve ser proposta exclusivamente pelo(s) aluno(s) com base nos conhecimentos adquiridos nas aulas. Consulte as notas de aula e o livro texto quando necessário.

Programa

Descrição Geral. Implemente um programa em C por meio do qual o usuário possa inserir, remover e atualizar dados de arquivos binários, bem como criar um índice para indexar esses arquivos. O índice é caracterizado por ser do tipo índices simples.

Importante. A definição da sintaxe de cada comando bem como sua saída devem seguir estritamente as especificações definidas em cada funcionalidade. Para especificar a sintaxe de execução, considere que o programa seja chamado de “programaTrab”. Essas orientações devem ser seguidas uma vez que a correção do funcionamento do programa

se dará de forma automática. De forma geral, a primeira entrada da entrada padrão é sempre o identificador de suas funcionalidades, conforme especificado a seguir.

Modularização. É importante modularizar o código. Trechos de programa que aparecerem várias vezes devem ser modularizados em funções e procedimentos.

Descrição Específica. O programa deve oferecer as seguintes funcionalidades:

Na linguagem SQL, o comando CREATE TABLE é usado para criar uma tabela, a qual é implementada como um arquivo. Geralmente, indica-se um campo (ou um conjunto de campos) que consiste na chave primária da tabela. Isso é realizado especificando-se a cláusula PRIMARY KEY. A funcionalidade [4] representa um exemplo de implementação de um índice simples definido sobre o campo chave primária de um arquivo de dados.

Na linguagem SQL, o comando CREATE INDEX é usado para criar um índice sobre um campo (ou um conjunto de campos) de busca. A funcionalidade [4] representa um exemplo de implementação de um índice simples definido sobre o campo chave primária de um arquivo de dados.

[4] Crie um arquivo de índice simples primário sobre um arquivo de dados de entrada. O campo de busca é o *id*, o qual identifica univocamente cada registro do arquivo de dados e não aceita valores repetidos e nem valores nulos. A criação deve ser feita considerando que o arquivo de dados já existe e já possui dados.

O arquivo de índice é definido da seguinte forma:

Registro de Cabeçalho. O registro de cabeçalho deve conter o seguinte campo:

- *status*: indica a consistência do arquivo de dados, devido à queda de energia, travamento do programa, etc. Pode assumir os valores '0', para indicar que o arquivo de índice está inconsistente, ou '1', para indicar que o arquivo de índice está consistente. Ao se abrir um arquivo para escrita, seu *status* deve ser '0' e, ao finalizar o uso desse arquivo, seu *status* deve ser '1' – tamanho: *string* de 1 byte.

Registros de Dados. Os registros de dados são de tamanho fixo, com campos de tamanho fixo, definidos da seguinte forma:

- *id*: chave de busca referente ao código identificador – inteiro – tamanho: 4 bytes.
- *byteoffset*: *byte offset* do registro do arquivo de dados referente ao *id* – inteiro – tamanho: 8 bytes.

Antes de terminar a execução da funcionalidade, deve ser utilizada a função `binarioNaTela`, disponibilizada na página do projeto da disciplina, para mostrar a saída do arquivo de índice.

Entrada do programa para a funcionalidade [4]:

```
4 arquivoDados.bin arquivoIndice.bin
```

onde:

- `arquivoDados.bin` é o arquivo binário que contém dados de entrada e que foi gerado conforme as especificações descritas nos trabalhos práticos.
- `arquivoIndice.bin` é um arquivo binário que indexa o arquivo de dados `arquivoDados.bin` e que é gerado conforme as especificações descritas neste trabalho prático.

Saída caso o programa seja executado com sucesso:

Listar o arquivo de índice no formato binário usando a função fornecida `binarioNaTela`.

Mensagem de saída caso algum erro seja encontrado:

Falha no processamento do arquivo.

Exemplo de execução:

```
./programaTrab  
4 jogador.bin jogadorIndice.bin
```

usar a função `binarioNaTela` antes de terminar a execução da funcionalidade, para mostrar a saída do arquivo `jogadorIndice.bin`.

Na linguagem SQL, o comando DELETE é usado para remover dados em uma tabela. Para tanto, devem ser especificados quais dados (ou seja, registros) devem ser removidos, de acordo com algum critério.

DELETE FROM tabela (ou seja, arquivo que contém os campos)

WHERE critério de seleção (ou seja, critério de busca)

A funcionalidade [5] representa um exemplo de implementação do comando DELETE.

[5] Permita a remoção lógica de registros de um arquivo de dados de entrada, baseado na *abordagem dinâmica* de reaproveitamento de espaços de registros logicamente removidos. A implementação dessa funcionalidade deve ser realizada usando o conceito de *lista de registros logicamente removidos*, e deve seguir estritamente a matéria apresentada em sala de aula. A lista deve ser implementada como uma lista ordenada de forma crescente considerando os tamanhos dos registros logicamente removidos e deve ser aplicada a estratégia Best Fit. Os registros a serem removidos devem ser aqueles que satisfaçam um critério de busca determinado pelo usuário. Se o critério de busca for definido em termos do campo *id* (ou qualquer combinação envolvendo esse campo) a busca deve ser realizada usando o índice criado na funcionalidade [4]. Caso contrário, a busca deve ser realizada conforme a especificação da funcionalidade [3]. Note que qualquer campo pode ser utilizado como forma de remoção. Ao se remover um registro, os valores dos *bytes* referentes aos campos já armazenados no registro devem permanecer os mesmos, com exceção dos valores dos campos relacionados ao tratamento da lista encadeada. A funcionalidade [5] deve ser executada n vezes seguidas. Em situações nas quais um determinado critério de busca não seja satisfeito, ou seja, caso a solicitação do usuário não retorne nenhum registro a ser removido, o programa deve continuar a executar as remoções até completar as n vezes seguidas. Antes de terminar a execução da funcionalidade, deve ser utilizada a função `binarioNaTela`, disponibilizada na página do projeto da disciplina, para mostrar a saída dos arquivos binários (de dados e de índice).

Entrada do programa para a funcionalidade [5]:

```
5 arquivoDados.bin arquivoIndice.bin n
```

```
x1 nomeCampo1 valorCampo1 ... nomeCampox1 valorCampox1  
x2 nomeCampo1 valorCampo1 ... nomeCampox2 valorCampox2  
...  
xn nomeCampo1 valorCampo1 ... nomeCampox2 valorCampoxn
```

onde:

- `arquivoDados.bin` é o arquivo binário que contém dados dos jogadores e que foi gerado conforme as especificações descritas no trabalho prático introdutório. As remoções a serem realizadas nessa funcionalidade devem ser feitas nesse arquivo.
- `arquivoIndice.bin` é um arquivo binário que indexa o arquivo de dados `arquivoDados.bin` e que foi gerado conforme as especificações descritas neste trabalho prático. As remoções realizadas no arquivo de dados devem ser refletidas nesse arquivo.
- `n` é o número de remoções a serem realizadas. Para cada remoção, deve ser informado em uma linha diferente o número de campos `x` usados como critério de busca e, para cada campo usado como critério de busca, o nome do campo e o seu respectivo valor do campo. Deve ser deixado um espaço em branco entre o nome do campo e o valor do campo. Os valores dos campos do tipo *string* devem ser especificados entre aspas duplas ("").

Saída caso o programa seja executado com sucesso:

Listar o arquivo de dados e o arquivo de índice no formato binário usando a função fornecida `binarioNaTela`.

Mensagem de saída caso algum erro seja encontrado:

Falha no processamento do arquivo.

Exemplo de execução:

```
./programaTrab  
5 jogador.bin jogadorIndice.bin 2  
1 id 261529  
2 nacionalidade "SPAIN" idade 24
```

usar a função `binarioNaTela` antes de terminar a execução da funcionalidade, para mostrar a saída dos arquivos `jogador.bin` e `jogadorIndice.bin`, os quais foram atualizados frente às remoções.

Na linguagem SQL, o comando INSERT INTO é usado para inserir dados em uma tabela. Para tanto, devem ser especificados os valores a serem armazenados em cada coluna da tabela, de acordo com o tipo de dado definido. A funcionalidade [6] representa exemplo de implementação do comando INSERT INTO.

[6] Permita a inserção de novos registros em um arquivo de dados de entrada, baseado na *abordagem dinâmica* de reaproveitamento de espaços de registros logicamente removidos. A implementação dessa funcionalidade deve ser realizada usando o conceito de *lista de registros logicamente removidos*, e deve seguir estritamente a matéria apresentada em sala de aula. A lista deve ser implementada como uma lista ordenada de forma crescente considerando os tamanhos dos registros logicamente removidos e deve ser aplicada a estratégia Best Fit. O lixo que permanece no registro logicamente removido e que é reutilizado deve ser identificado pelo caractere '\$'. Na entrada desta funcionalidade, os dados são referentes aos seguintes campos, na seguinte ordem: id, idade, nomeJogador, nacionalidade, nomeClube. Campos com valores nulos, na entrada da funcionalidade, devem ser identificados com NULO. Os valores dos campos do tipo *string* devem ser especificados entre aspas duplas ("). Para a manipulação de *strings* com aspas duplas, pode-se usar a função scan_quote_string disponibilizada na página do projeto da disciplina. A funcionalidade [6] deve ser executada *n* vezes seguidas. Antes de terminar a execução da funcionalidade, deve ser utilizada a função binarioNaTela, disponibilizada na página do projeto da disciplina, para mostrar a saída dos arquivos binários (de dados e de índice).

Entrada do programa para a funcionalidade [6]:

```
6 arquivoDados.bin arquivoIndice.bin n
id1, idade1, nomeJogador1, nacionalidade1, nomeClube1
id2, idade2, nomeJogador2, nacionalidade2, nomeClube2
...
idn, idaden, nomeJogadorn, nacionalidaden, nomeCluben
```

onde:

- arquivoDados.bin é o arquivo binário que contém dados dos jogadores e que foi gerado conforme as especificações descritas no trabalho prático introdutório. As inserções a serem realizadas nessa funcionalidade devem ser feitas nesse arquivo.
- arquivoIndice.bin é um arquivo binário que indexa o arquivo de dados arquivoDados.bin e que foi gerado conforme as especificações descritas neste trabalho prático. As inserções realizadas no arquivo de dados devem ser refletidas nesse arquivo.
- n é o número de inserções a serem realizadas. Para cada inserção, deve ser informado os valores a serem inseridos no arquivo, considerando os seguintes campos, na seguinte ordem: id, idade, nomeJogador, nacionalidade, nomeClube. Valores nulos devem ser identificados, na entrada da funcionalidade, por NULO. Cada uma das n inserções deve ser especificada em uma linha diferente. Deve ser deixado um espaço em branco entre os valores dos campos. Os valores dos campos do tipo *string* devem ser especificados entre aspas duplas (").

Saída caso o programa seja executado com sucesso:

Listar o arquivo de dados e o arquivo de índice no formato binário usando a função fornecida binarioNaTela.

Mensagem de saída caso algum erro seja encontrado:

Falha no processamento do arquivo.

Exemplo de execução:

```
./programaTrab
6 jogador.bin jogadorIndice.bin 2
261529 24 "D. FOMIN" "RUSSIA" NULO
247106 21 "A. BERNABEI" "ARGENTINA" CELTIC
```

usar a função binarioNaTela antes de terminar a execução da funcionalidade, para mostrar a saída dos arquivos jogador.bin e jogadorIndice.bin, os quais foram atualizados frente às inserções.

Restrições

As seguintes restrições têm que ser garantidas no desenvolvimento do trabalho.

[1] O arquivo de dados e o arquivo de índice devem ser gravados em disco no **modo binário**. O modo texto não pode ser usado.

[2] Os dados do registro descrevem os nomes dos campos, os quais não podem ser alterados. Ademais, todos os campos devem estar presentes na implementação, e nenhum campo adicional pode ser incluído. O tamanho e a ordem de cada campo deve obrigatoriamente seguir a especificação.

[3] Deve haver a manipulação de valores nulos, conforme as instruções definidas.

[4] Não é necessário realizar o tratamento de truncamento de dados.

[5] Devem ser exibidos avisos ou mensagens de erro de acordo com a especificação de cada funcionalidade.

[6] Os dados devem ser obrigatoriamente escritos campo a campo. Ou seja, não é possível escrever os dados registro a registro. Essa restrição refere-se à entrada/saída, ou seja, à forma como os dados são escritos no arquivo.

[7] O(s) aluno(s) que desenvolveu(desenvolveram) o trabalho prático deve(m) constar como comentário no início do código (i.e. NUSP e nome do aluno). Para trabalhos desenvolvidos por mais do que um aluno, não será atribuída nota ao aluno cujos dados não constarem no código fonte.

[8] Todo código fonte deve ser documentado. A **documentação interna** inclui, dentre outros, a documentação de procedimentos, de funções, de variáveis, de partes do código

fonte que realizam tarefas específicas. Ou seja, o código fonte deve ser documentado tanto em nível de rotinas quanto em nível de variáveis e blocos funcionais.

[9] A implementação deve ser realizada usando a linguagem de programação C. As funções das bibliotecas <stdio.h> devem ser utilizadas para operações relacionadas à escrita e leitura dos arquivos. A implementação não pode ser feita em qualquer outra linguagem de programação. O programa executará no [run.codes].

Material para Entregar

Arquivo compactado. Deve ser preparado um arquivo .zip contendo:

- Código fonte do programa devidamente documentado.
- Makefile para a compilação do programa.
- Um vídeo gravado pelos integrantes do grupo, o qual deve ter, no máximo, 5 minutos de gravação. O vídeo deve explicar o trabalho desenvolvido. Ou seja, o grupo deve apresentar: cada funcionalidade e uma breve descrição de como a funcionalidade foi implementada. Todos os integrantes do grupo devem participar do vídeo, sendo que o tempo de apresentação dos integrantes deve ser balanceado. Ou seja, o tempo de participação de cada integrante deve ser aproximadamente o mesmo. O uso da webcam é obrigatório.

Instruções para fazer o arquivo makefile. No [run.codes] tem uma orientação para que, no makefile, a diretiva “all” contenha apenas o comando para compilar seu programa e, na diretiva “run”, apenas o comando para executá-lo. Assim, a forma mais simples de se fazer o arquivo makefile é:

```
all:
    gcc -o programaTrab *.c
run:
    ./programaTrab
```

Lembrando que *.c já engloba todos os arquivos .c presentes no seu zip. Adicionalmente, no arquivo Makefile é importante se ter um *tab* nos locais colocados acima, senão ele pode não funcionar.

Instruções de entrega.

O programa deve ser submetido via [run.codes]:

- página: <https://run.codes/Users/login>
- **Turma 1** (segunda-feira): código de matrícula: **SQRG**
- **Turma 2** (terça-feira): código de matrícula: **DPZ4**

O vídeo gravado deve ser submetido por meio da página da disciplina no e-disciplinas, no qual o grupo vai informar o nome de cada integrante, o número do grupo e um link que contém o vídeo gravado. Ao submeter o link, verifique se o mesmo pode ser acessado. Vídeos cujos links não puderem ser acessados receberão nota zero.

Critério de Correção

Critério de avaliação do trabalho. Na correção do trabalho, serão ponderados os seguintes aspectos.

- Corretude da execução do programa.
- Atendimento às especificações do registro de cabeçalho e dos registros de dados.
- Atendimento às especificações da sintaxe dos comandos de cada funcionalidade e do formato de saída da execução de cada funcionalidade.
- Qualidade da documentação entregue. A documentação interna terá um peso considerável no trabalho.
- Vídeo. Integrantes que não participarem da apresentação receberão nota 0 no trabalho correspondente.

Casos de teste no [run.codes]. Juntamente com a especificação do trabalho, serão disponibilizados 70% dos casos de teste no [run.codes], para que os alunos possam

avaliar o programa sendo desenvolvido. Os 30% restantes dos casos de teste serão utilizados nas correções.

Restrições adicionais sobre o critério de correção.

- A não execução de um programa devido a erros de compilação implica que a nota final da parte do trabalho será igual a zero (0).
- O não atendimento às especificações do registro de cabeçalho e dos registros de dados implica que haverá uma diminuição expressiva na nota do trabalho.
- O não atendimento às especificações de sintaxe dos comandos de cada funcionalidade e do formato de saída da execução de cada funcionalidade implica que haverá uma diminuição expressiva na nota do trabalho.
- A ausência da documentação implica que haverá uma diminuição expressiva na nota do trabalho.
- A realização do trabalho prático com alunos de turmas diferentes implica que haverá uma diminuição expressiva na nota do trabalho.
- A inserção de palavras ofensivas nos arquivos e em qualquer outro material entregue implica que a nota final da parte do trabalho será igual a zero (0).
- Em caso de plágio, as notas dos trabalhos envolvidos serão zero (0).

Bom Trabalho!