

INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

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

# Docente Profa. Dra. Cristina Dutra de Aguiar

cdac@icmc.usp.br

#### **Monitores**

### Caio Capocasali da Silva

caio.capocasali@usp.br ou telegram: @CaioCapocasali

Renan Banci Catarin

renanbcatarin@usp.br ou telegram: @Reckat

Renan Trofino Silva

renan.trofino@usp.br ou telegram: @renan823

#### Trabalho Introdutório

Este trabalho tem como objetivo obter dados de um arquivo de entrada e gerar um arquivo binário com eles, bem como realizar operações de busca com e sem um índice primário. Ele é um trabalho introdutório, de forma que será usado como base para o desenvolvimento dos demais trabalhos da disciplina.

O trabalho deve ser feito por 2 alunos. A solução deve ser proposta exclusivamente pelos alunos com base nos conhecimentos adquiridos nas aulas. Consulte as notas de aula e o livro texto quando necessário.

### Fundamentos da disciplina de Bases de Dados

A disciplina de Organização de Arquivos é uma disciplina fundamental para a disciplina de Bases de Dados. A definição dos trabalhos práticos é feita considerando esse aspecto, ou seja, os trabalhos são especificados em termos de várias funcionalidades, e essas funcionalidades são relacionadas tanto com desafios enfrentados no mercado de trabalho quanto com as funcionalidades da linguagem SQL (Structured Query Language), que é a linguagem utilizada por sistemas gerenciadores de banco de dados (SGBDs) relacionais. As características e o detalhamento de SQL serão aprendidos na disciplina de Bases de Dados. Porém, por meio do desenvolvimento deste trabalho prático, os alunos podem entrar em contato com alguns comandos da linguagem SQL e verificar como eles são implementados.



NSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

Os trabalhos práticos têm como objetivo armazenar e recuperar dados relacionados às *pessoas que seguem pessoas em uma rede social*. Na disciplina de Bases de Dados, é ensinado que o projeto deve ser feito em dois arquivos. O primeiro arquivo é relacionado às pessoas, armazenando apenas dados relacionados a essas pessoas. O segundo arquivo é um arquivo de relacionamento, que relaciona pessoas que seguem outras pessoas. Visando atender aos requisitos de um bom projeto do banco de dados, são definidos dois arquivos de dados a serem utilizados nos trabalhos práticos: arquivo de dados **pessoa** e arquivo de dados **segue**. Neste primeiro trabalho prático, são implementados aspectos de criação e busca do arquivo **pessoa**.

### Descrição do Arquivo Pessoa

O arquivo de dados **pessoa** possui um registro de cabeçalho e 0 ou mais registros de dados, conforme a definição a seguir.

Registro de Cabeçalho. O registro de cabeçalho deve conter os seguintes campos:

- *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 dados está inconsistente, ou '1', para indicar que o arquivo de dados 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.
- quantidadePessoas: armazena a quantidade de pessoas (ou o número de registros que não foram removidos) presentes no arquivo. Ou seja, esse valor não inclui os registros logicamente marcados como removidos. Deve ser iniciado com o valor '0' e deve ser incrementado e decrementado quando necessário – tamanho: inteiro de 4 bytes.
- quantidadeRemovidos: armazena a quantidade de pessoas que foram removidas
   (ou o número de registros logicamente removidos) presentes no arquivo. Deve
   ser iniciado com o valor '0' e deve ser incrementado e decrementado quando
   necessário tamanho: inteiro de 4 bytes.





NSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

 proxByteOffset: armazena o valor do próximo byte offset disponível. Deve ser iniciado com o valor '17' e deve ser alterado sempre que necessário – tamanho: inteiro de 8 bytes.

Representação Gráfica do Registro de Cabeçalho. O tamanho do registro de cabeçalho deve ser de 17 bytes, representado da seguinte forma:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
status	quantidadePessoas				quantidadeRemovidos					proxByteOffset						

### Observações Importantes.

- O registro de cabeçalho deve seguir estritamente a ordem definida na sua representação gráfica.
- Os campos são de tamanho fixo. Portanto, os valores alfanuméricos que forem armazenados não devem ser finalizados por '\0'.
- Neste projeto, o conceito de página de disco não está sendo considerado.

**Registros de Dados.** Os registros de dados são de tamanho variável, com campos de tamanho fixo e campos de tamanho variável. Para os campos e os registros de tamanho variável, deve ser usado o método de indicador de tamanho.

Os campos de tamanho fixo são definidos da seguinte forma:

- *idPessoa*: código sequencial que identifica univocamente cada registro do arquivo de dados inteiro tamanho: 4 bytes.
- *idadePessoa*: idade da pessoa inteiro tamanho: 4 bytes.

Os campos de tamanho variável são definidos da seguinte forma:

- *nomePessoa*: nome completo da pessoa *string*
- nomeUsuario: nome da pessoa na rede social string

Adicionalmente, os seguintes campos de tamanho fixo também compõem cada registro. Esses campos são necessários para o gerenciamento de registros logicamente removidos e para o controle dos campos e registros de tamanho variável.

• removido: indica se o registro está logicamente removido. Pode assumir os valores '1', para indicar que o registro está marcado como logicamente



NSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

removido, ou '0', para indicar que o registro não está marcado como removido.

- tamanho: *string* de 1 byte.
- *tamanhoRegistro*: número de bytes do registro inteiro tamanho: 4 bytes.
- tamanhoNomePessoa: número de bytes do campo nomePessoa inteiro tamanho: 4 bytes.
- tamanhoNomeUsuario: número de bytes do campo nomeUsuario inteiro tamanho: 4 bytes.

Os dados são fornecidos juntamente com a especificação deste trabalho prático por meio de um arquivo .csv, sendo que sua especificação está disponível na página da disciplina. No arquivo .csv, o separador de campos é vírgula (,) e o primeiro registro especifica o que cada coluna representa (ou seja, contém a descrição do conteúdo das colunas). Adicionalmente, campos nulos são representados por espaço em branco.

Representação Gráfica dos Registros de Dados. Cada registro de dados deve ser representado da seguinte forma:

0	1	2	3		4	5	6	7	8	9	10	11	12	13	14
remo		tam	anhoR	egistro			idPes	ssoa		idadeF	tamanho				
vido														Nome	Pessoa
15	16	17													
	nomePessoa (variável)				tamanho NomeUsuario				nomeUsuario (variável)						

#### Observações Importantes.

- Cada registro de dados deve seguir estritamente a ordem definida na sua representação gráfica.
- Os valores alfanuméricos dos campos de tamanhão variável não devem ser finalizados por '\0', desde que o tamanho deles são armazenados no registro.
- O campo idPessoa não pode assumir valores nulos ou valores repetidos. O arquivo .csv com os dados de entrada já garante essas características.
- O campo nomeUsuario n\u00e3o pode assumir valores nulos ou valores repetidos. O
  arquivo .csv com os dados de entrada j\u00e1 garante essas caracter\u00edsticas.
- Os valores nulos nos campos de tamanho fixo devem ser manipulados da seguinte forma. Os valores nulos devem ser representados pelo valor -1 quando





INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

forem inteiros ou devem ser totalmente preenchidos pelo lixo '\$' quando forem do tipo *string*.

- Os valores nulos nos campos de tamanho variável devem ser manipulados da seguinte forma. Apenas o indicador de tamanho deve ser armazenado, com o valor zero.
- Deve ser feita a diferenciação entre o espaço utilizado e o lixo. Sempre que houver lixo, ele deve ser identificado pelo caractere '\$'. Nenhum byte do registro deve permanecer vazio, ou seja, cada byte deve armazenar um valor válido ou '\$'.
- Não existe a necessidade de truncamento dos dados. O arquivo .csv com os dados de entrada já garante essa característica.
- Neste projeto, o conceito de página de disco não está sendo considerado.

### Descrição do Arquivo de Índice Primário

O arquivo de dados **pessoa** é indexado por um índice primário, que é definido sobre o campo *idPessoa*. O índice primário também é um arquivo, sendo chamado de arquivo de índice **indexaPessoa**. A especificação de **indexaPessoa** é feita a seguir.

Registro de Cabeçalho. O registro de cabeçalho deve conter os seguintes campos:

• *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 dados está inconsistente, ou '1', para indicar que o arquivo de dados 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.

Representação Gráfica do Registro de Cabeçalho. O tamanho do registro de cabeçalho deve ser de 12 bytes, representado da seguinte forma:





INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

### Observações Importantes.

- O registro de cabeçalho deve seguir estritamente a ordem definida na sua representação gráfica.
- Para caber em páginas de disco (que são definidas em potência de 2), o registro de cabeçalho foi definido como uma potência de 2, no mesmo tamanho dos registros de dados). Portanto, o registro de cabeçalho tem o tamanho de 12 bytes, sendo 1 byte preenchido com dados necessários para o desenvolvimento do trabalho, e os bytes restantes preenchidos com lixo. O lixo é representado pelo caractere '\$'.

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

- *idPessoa*: código sequencial que identifica univocamente cada registro de pessoa armazenado no arquivo de dados pessoa: *inteiro tamanho: 4 bytes*.
- byteOffset: byte offset do registro do arquivo de dados que se refere ao idPessoa
   inteiro tamanho: 8 bytes.

Representação Gráfica dos Registros de Dados. O tamanho de cada registro de dados deve ser de 8 bytes, representado da seguinte forma:

0	1	2	3	4	5	6	7	8	9	10	11
	idPe	ssoa				i	byteOff	set			

#### Observações Importantes.

- Cada registro de dados deve seguir estritamente a ordem definida na sua representação gráfica.
- Os campos idPessoa e byteOffset não aceitam valores nulos.
- Os valores de *idPessoa* devem ser ordenados de forma crescente.
- Neste projeto, o conceito de página de disco não está sendo considerado.



INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

### Programa

**Descrição Geral**. Implemente um programa em C por meio do qual o usuário possa obter dados de um arquivo de entrada e gerar arquivos binários com esses dados, bem como realizar operações de busca nesses arquivos binários com e sem o uso de um índice primário.

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:



INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

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 [1] representa um exemplo de implementação da criação de um índice primário definido sobre o campo chave primária de um arquivo de dados.

[1] Crie um arquivo de índice primário para um arquivo de dados de entrada. O arquivo de índice primário deve ser gerado de acordo com as especificações da **Descrição do Arquivo de Índice Primário** definidas neste trabalho prático. No momento da criação, o arquivo de índice primário apenas possui o registro de cabeçalho.

### Entrada do programa para a funcionalidade [1]:

1 arquivoIndicePrimario.bin

#### onde:

- arquivoIndicePrimario.bin é um arquivo de índice primário no formato binário, o qual é gerado de acordo com as especificações da **Descrição do Arquivo de Índice Primário** definidas neste trabalho prático.

### Saída caso o programa seja executado com sucesso:

Listar todos os registros do arquivo de índice primário 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
- 1 indexaPessoa.bin

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



INSTITUTO DE CIENCIAS MATEMATICAS E DE COMPUTAÇAC Departamento de Ciências de Computação

Na linguagem SQL, o comando CREATE TABLE é usado para criar uma tabela, a qual é implementada como um arquivo. Geralmente, uma tabela possui um nome (que corresponde ao nome do arquivo) e várias colunas, as quais correspondem aos campos dos registros do arquivo de dados. A funcionalidade [2] representa um exemplo de implementação do comando CREATE TABLE.

[2] Permita a leitura de vários registros obtidos a partir de um arquivo de entrada no formato esv e a gravação desses registros em um arquivo de dados de saída. O arquivo de entrada no formato esv é fornecido juntamente com a especificação do projeto, enquanto o arquivo de dados de saída (arquivo **pessoa**) deve ser gerado de acordo com as especificações deste trabalho prático. Nesta funcionalidade, ocorrem várias inserções no arquivo **pessoa**. O arquivo de índice **indexaPessoa** já teve estar criado. Assim, para cada inserção de um registro no arquivo de dados pessoa, deve ser inserido também uma entrada correspondente no arquivo de índice **indexaPessoa**. Antes de terminar a execução da funcionalidade, deve ser utilizada a função binarioNaTela, disponibilizada na página do projeto da disciplina, para mostrar os dados do arquivo **pessoa** e do arquivo **indexaPessoa**. A função binarioNaTela deve ser usada depois que o arquivo é fechado e antes de terminar a execução da funcionalidade.





TITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO

Departamento de Ciências de Computação

### Entrada do programa para a funcionalidade [2]:

2 arquivoEntrada.csv arquivoSaida.bin arquivoIndicePrimario.bin onde:

- arquivoEntrada.csv é um arquivo .csv que contém os valores dos campos dos registros a serem armazenados no arquivo de dados **pessoa**.
- arquivosaida.bin é o arquivo de dados **pessoa** no formato binário, o qual é gerado de acordo com as especificações da **Descrição do Arquivo de Índice Primário** definidas neste trabalho prático.
- arquivoIndicePrimario.bin é um arquivo de índice primário no formato binário, o qual é gerado de acordo com as especificações da **Descrição do Arquivo de Dados Pessoa** definidas neste trabalho prático.

#### 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
- 2 pessoa.csv pessoa.bin indexaPessoa.bin

usar a função binarioNaTela antes de terminar a execução da funcionalidade, para mostrar a saída dos arquivos binários pessoa.bin e indexaPessoa.bin.





INSTITUTO DE CIENCIAS MATEMATICAS E DE COMPUTAÇAC Departamento de Ciências de Computação

Na linguagem SQL, o comando SELECT é usado para listar os dados de uma tabela. Existem várias cláusulas que compõem o comando SELECT. O comando mais básico consiste em especificar as cláusulas SELECT e FROM, da seguinte forma:

SELECT lista de colunas (ou seja, campos a serem exibidos na resposta)

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

A funcionalidade [3] representa um exemplo de implementação do comando SELECT. Como todos os registros devem ser recuperados nessa funcionalidade, sua implementação consiste em percorrer sequencialmente o arquivo.

[3] Permita a recuperação dos dados de todos os registros armazenados no arquivo de dados **pessoa**, mostrando os dados de forma organizada na saída padrão para permitir a distinção dos campos e registros. O tratamento de 'lixo' deve ser feito de forma a permitir a exibição apropriada dos dados. Registros marcados como logicamente removidos não devem ser exibidos.





INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

### Entrada do programa para a funcionalidade [3]:

3 arquivoEntrada.bin

#### onde:

- arquivoEntrada.bin é o arquivo de dados **pessoa** no formato binário, o qual é gerado de acordo com as especificações da **Descrição do Arquivo Pessoa** definidas neste trabalho prático.

#### Saída caso o programa seja executado com sucesso:

A saída deve ser exibida da seguinte forma. Para cada registro:

"Dados da pessoa de codigo " escrever o valor de idPessoa

"Nome: " escrever o valor de nomePessoa

"Idade: " escrever o valor de idadePessoa

"Usuario: " escrever o valor de nomeUsuario

Valores nulos devem substituídos pelo caractere "-".

Deve ser deixada uma linha em branco entre cada registro listado.

Ver exemplo ilustrado no exemplo de execução.

#### Mensagem de saída caso não existam registros:

Registro inexistente.

#### Mensagem de saída caso algum erro seja encontrado:

Falha no processamento do arquivo.

### Exemplo de execução (é mostrado um exemplo ilustrativo):

./programaTrab

3 pessoa.bin

Dados da pessoa de codigo 25 Nome: SAMANTHA PEREIRA SANTOS

Idade: 13

Usuario: SAMANTHAPS

linha em branco

Dados da pessoa de codigo 45 Nome: VITORIA PRADO CAMPOS

Idade: -

Usuario: VIVICA linha em branco

• • •



INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

Conforme visto na funcionalidade [3], na linguagem SQL o comando SELECT é usado para listar os dados de uma tabela. Existem várias cláusulas que compõem o comando SELECT. Além das cláusulas SELECT e FROM, outra cláusula muito comum é a cláusula WHERE, que permite que seja definido um critério de busca sobre um ou mais campos, o qual é nomeado como critério de seleção.

SELECT lista de colunas (ou seja, campos a serem exibidos na resposta)

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

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

A funcionalidade [4] representa um exemplo de implementação do comando SELECT considerando a cláusula WHERE. Como não existe índice definido sobre os campos dos registros, a implementação dessa funcionalidade consiste em percorrer sequencialmente o arquivo.

[4] Permita a recuperação dos dados de todos os registros do arquivo de dados **pessoa**, de forma que esses registros satisfaçam um critério de busca determinado pelo usuário. Qualquer campo pode ser utilizado como forma de busca. Porém, a busca não pode ser combinada. Ou seja, apenas um único campo deve ser utilizado como critério de busca. Por exemplo, é possível realizar a busca considerando somente o campo *idPessoa* ou somente o campo *idade*. Em situações nas quais a busca for feita considerando o campo *idPessoa*, deve ser utilizado o arquivo de índice **indexaPessoa** para se fazer a busca. Tem-se, nesse caso, uma busca indexada. Para os demais casos, deve ser feita uma busca sequencial. Esta funcionalidade pode retornar 0 registros (quando nenhum satisfaz ao critério de busca), 1 registro (quando apenas um satisfaz ao critério de busca), ou vários registros. Os valores dos campos do tipo *string* devem ser especificados entre aspas duplas (") na chamada da funcionalidade. Para a manipulação de *strings* com aspas duplas, pode-se usar a função scan\_quote\_string disponibilizada na página do projeto da disciplina. Registros marcados como logicamente removidos não devem ser exibidos.



INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

### Sintaxe do comando para a funcionalidade [4]:

- 4 arquivoEntrada.bin arquivoIndicePrimario.bin n
- 1 nomeCampo=valorCampo
- 2 nomeCampo=valorCampo

. . .

n nomeCampo<sub>1</sub>=valorCampo<sub>1</sub>

#### onde:

- arquivoEntrada.bin é o arquivo de dados **pessoa** no formato binário, o qual é gerado de acordo com as especificações da **Descrição do Arquivo Pessoa** definidas neste trabalho prático.
- arquivoIndicePrimario.bin é um arquivo de índice primário no formato binário, o qual é gerado de acordo com as especificações da **Descrição do Arquivo de Índice Primário** definidas neste trabalho prático.
- n é a quantidade de vezes que a busca deve acontecer.
- nome do Campo= valor do Campo referem-se ao nome e ao valor do campo que estão sendo procurados no arquivo de dados pessoa. Deve ser colocado um símbolo de igual (=) entre nomeCampo e valorCampo, sem espaço em branco. Os valores dos campos do tipo *string* devem ser especificados entre aspas duplas (").

#### Saída caso o programa seja executado com sucesso:

Podem ser encontrados vários registros que satisfaçam à condição de busca. A saída deve ser exibida da seguinte forma. Para cada registro:

"Dados da pessoa de codigo " escrever o valor de idPessoa

"Nome: " escrever o valor de nomePessoa

"Idade: " escrever o valor de idadePessoa

"Usuario: " escrever o valor de nomeUsuario

Valores nulos devem substituídos pelo caractere "-".

Deve ser deixada uma linha em branco entre cada registro listado.

Ver exemplo ilustrado no exemplo de execução.

Mensagem de saída caso não seja encontrado o registro que contém o valor do campo ou o campo pertence a um registro que esteja removido:

Registro inexistente.





INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

### Mensagem de saída caso algum erro seja encontrado:

Falha no processamento do arquivo.

### Exemplo de execução:

- ./programaTrab
- 4 pessoa.bin indexaPessoa.bin 1
- 1 idPessoa=25

Dados da pessoa de codigo 25 Nome: SAMANTHA PEREIRA CAMPOS

Idade: 13

Usuario: SAMANTHAPS

linha em branco

...

- ./programaTrab
- 4 pessoa.bin indexaPessoa.bin 2
- 1 idade=13

Dados da pessoa de codigo 25

Nome: CAMILA SANTOS

Idade: 13

Usuario: CAMILAS

linha em branco

•••

2 idPessoa=54

Dados da pessoa de codigo 54 Nome: MARIA DO NASCIMENTO

Idade: 37

Usuario: MNASCIMENTO

linha em branco

•••



INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

### Restrições

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

- [1] O arquivo de dados deve ser gravado 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



NSTITUTO DE CIENCIAS MATEMATICAS E DE COMPUTAÇAO Departamento de Ciências de Computação

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].

### Fundamentação Teórica

Conceitos e características dos diversos métodos para representar os conceitos de campo e de registro em um arquivo de dados podem ser encontrados nos *slides* de sala de aula e também no livro *File Structures* (*second edition*), de Michael J. Folk e Bill Zoellick.

#### 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.





NSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

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. Adicionalmente, para utilizar a função binarioNaTela, é necessário usar a flag -lmd. Assim, a forma mais simples de se fazer o arquivo makefile é:

all:

gcc -o programaTrab \*.c -lmd

run:

./programaTrab

Lembrando que \*.c já engloba todos os arquivos .c presentes no arquivo 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://runcodes.icmc.usp.br/

• Código de matrícula: **3YE6** 

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. Vídeos corrompidos ou que não puderem ser corretamente 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.



NSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO Departamento de Ciências de Computação

- 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!