# EDCO4B ESTRUTURAS DE DADOS 2

Aula 11 - Busca e Ordenação em Arquivos

Prof. Rafael G. Mantovani



## Licença

Este trabalho está licenciado com uma Licença CC BY-NC-ND 4.0:



Atribuição-NãoComercial-SemDerivações 4.0 Internacional (CC BY-NC-ND 4.0)

#### maiores informações:

https://creativecommons.org/licenses/by-nc-nd/4.0/deed.pt\_BR

## Roteiro

- 1 Introdução
- 2 Busca Binária
- **3** Key Sorting
- 4 Revisão
- 5 Referências

## Roteiro

- 1 Introdução
- 2 Busca Binária
- **3** Key Sorting
- 4 Revisão
- 5 Referências

O que ja fizemos:

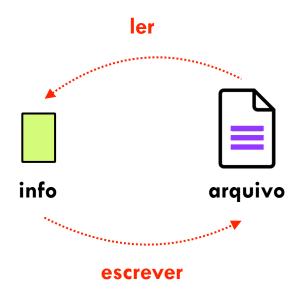
- O que ja fizemos:
  - como organizar informações em arquivos



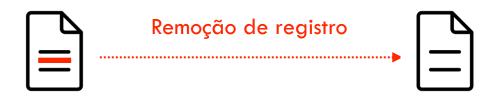
- O que ja fizemos:
  - como organizar informações em arquivos
  - diferentes formas de representar as informações nos arquivos



- O que ja fizemos:
  - como organizar informações em arquivos
  - diferentes formas de representar as informações nos arquivos
  - como ler e escrever informações dos arquivos



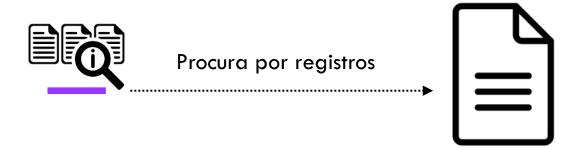
- O que ja fizemos:
  - como organizar informações em arquivos
  - diferentes formas de representar as informações nos arquivos
  - como ler e escrever informações dos arquivos
  - como remover registros de arquivos



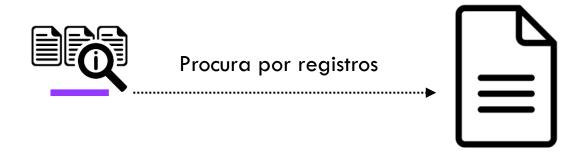
- O que ja fizemos:
  - como organizar informações em arquivos
  - diferentes formas de representar as informações nos arquivos
  - como ler e escrever informações dos arquivos
  - como remover registros de arquivos
  - como otimizar o uso dos arquivos / reuso



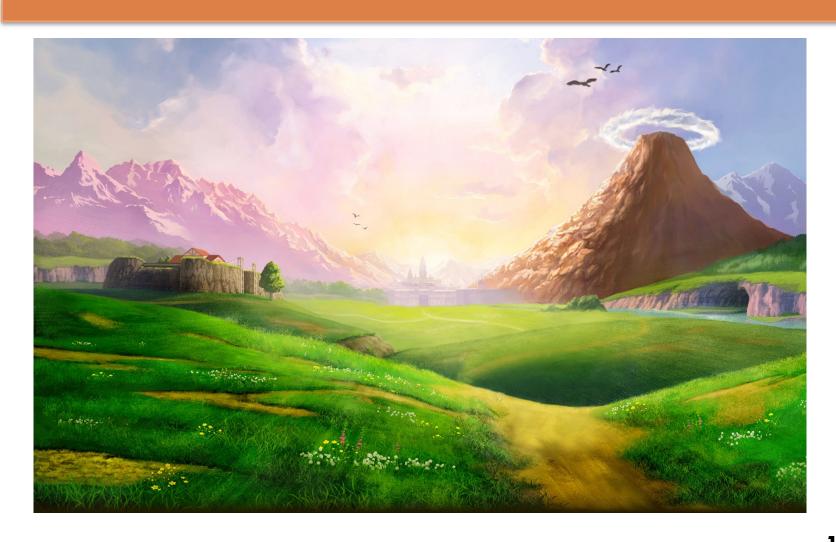
Como procuramos/recuperamos uma informação desejada?

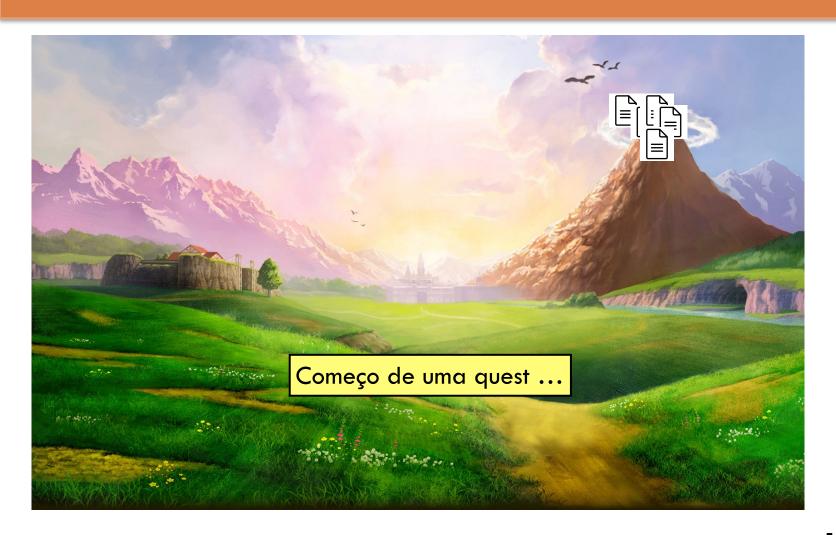


Como procuramos/recuperamos uma informação desejada?



- eficiência depende do número de leituras (seek)
- o que vale para procura também vale para ordenação





Até agora:



Procura por registros

Até agora:



Até agora:



Até agora:

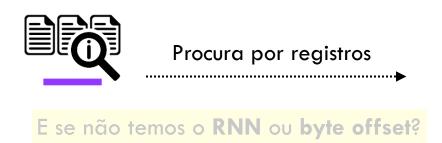


Procura por registros

E se não temos o RNN ou byte offset?



Até agora:





**Busca sequencial** 

Até agora:



Muito ruim! Precisamos de uma solução melhor!

## Roteiro

- 1 Introdução
- 2 Busca Binária
- 3 Key Sorting
- 4 Revisão
- 5 Referências



Procura por registros



registro tamanho fixo

```
Ames | Mary | 123 Maple | Stillwater | OK | 74075 | Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | Mason | Alan | 90 Eastgate | Ada | OK | 74820 | Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 | Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 |
```

#### Para funcionar:

## Key (chave canônica)

```
Ames | Mary | 123 Maple | Stillwater | OK | 74075 | ...
Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | ...
Mason | Alan | 90 Eastgate | Ada | OK | 74820 | ...
Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | ...
Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | ...
DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana | PR | 86800 | ...
Da Salsicha | Tamara | 10 Compila Street | Londrina | PR | 85800 | ...
Vegano | Fernando | 07 Redes | N7 | Apucarana | PR | 86800 | ...
Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | ...
Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | ...
```

## Key (chave canônica)

MARY AMES
SEBASTIAN MORISSON
ALAN MASON
MARTHA BROWN
RAFAEL MANTOVANI
ANDREZÃO DAMASSA
TAMARA DA SALSICHA
FERNANDO VEGANO
MAURICIO MAUMAU
ERI NALDO

Ames | Mary | 123 Maple | Stillwater | OK | 74075 |

Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 |

Mason | Alan | 90 Eastgate | Ada | OK | 74820 |

Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 |

Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 |

DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 |

Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 |

Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 |

Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 |

Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 |

#### Arquivo Ordenado em Ordem Crescente

ANDREZÃO DAMASSA
ALAN MASON
ERI NALDO
MARY AMES
FERNANDO VEGANO
MARTHA BROWN
MAURICIO MAUMAU
RAFAEL MANTOVANI
SEBASTIAN MORISSON
TAMARA DA SALSICHA

```
DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 | ...

Mason | Alan | 90 Eastgate | Ada | OK | 74820 | ...

Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | ...

Ames | Mary | 123 Maple | Stillwater | OK | 74075 | ...

Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | ...

Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | ...

Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | ...

Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | ...

Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | ...

Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | ...
```

```
DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 | ... Mason | Alan | 90 Eastgate | Ada | OK | 74820 | ... Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | ... Ames | Mary | 123 Maple | Stillwater | OK | 74075 | ... Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | ... Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | ... Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | ... Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | ... Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | ... Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | ...
```

#### **Procurar:**

| Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 |

```
DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 | ... Mason | Alan | 90 Eastgate | Ada | OK | 74820 | ... Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | ... Ames | Mary | 123 Maple | Stillwater | OK | 74075 | ... Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | ... Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | ... Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | ... Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | ... Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | ... Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | ...
```

#### **Procurar:**

| Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 |

#### **RRNs**

```
DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 | ...

Mason | Alan | 90 Eastgate | Ada | OK | 74820 | ...

Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | ...

Ames | Mary | 123 Maple | Stillwater | OK | 74075 | ...

Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | ...

Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | ...

Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | ...

Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | ...

Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | ...

Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | ...
```

#### **Procurar:**

| Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 |

#### **RRNs**

```
Início → 0

DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 | ...

Mason | Alan | 90 Eastgate | Ada | OK | 74820 | ...

Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | ...

Ames | Mary | 123 Maple | Stillwater | OK | 74075 | ...

Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | ...

Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | ...

Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | ...

Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | ...

Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | ...

Fim → 9

Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | ...
```

Meio = 
$$(Início + Fim)/2 = (0 + 10)/2 = 5$$

```
        Procurar:
        Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 |

        RRNs

        Início → 0
        DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana | PR | 86800 | ...

        1
        Mason | Alan | 90 Eastgate | Ada | OK | 74820 | ...
        ...

        2
        Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | ...
        ...

        3
        Ames | Mary | 123 Maple | Stillwater | OK | 74075 | ...
        ...

        Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | ...
        ...

        5
        Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | ...

        6
        Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | ...

        7
        Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | ...

        8
        Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | ...

        Pim → 9
        Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | ...
```

Meio = (Início + Fim)/2 = (0 + 9)/2 = 4.5 = 4

#### Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | Procurar: **RRNs** DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 | ... Mason | Alan | 90 Eastgate | Ada | OK | 74820 | ..... Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | ..... Ames | Mary | 123 Maple | Stillwater | OK | 74075 | ..... Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | ..... Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | .... Início → Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | ..... Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | ..... Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | ..... Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | ..... Fim

# Procurar: | Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | RRNs DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 | ... | Mason | Alan | 90 Eastgate | Ada | OK | 74820 | ... | Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | ... | Ames | Mary | 123 Maple | Stillwater | OK | 74075 | ... | Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | ... | Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | ... | Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | ... | Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | ... | Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | ... | Fim → 9 Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | ...

Meio = (Início + Fim)/2 = (5 + 9)/2 = 7

```
Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 |
   Procurar:
        RRNs
              DaMassa | Andrezão | 67 Arquitetura dos computa | Apucarana PR | 86800 | ...
              Mason | Alan | 90 Eastgate | Ada | OK | 74820 | .....
              Naldo | Eri | 12 FPGA | Apucarana | PR | 86800 | .....
              Ames | Mary | 123 Maple | Stillwater | OK | 74075 | .....
              Vegano | Fernando | 07 Redes N7 | Apucarana | PR | 86800 | .....
              Brown | Martha | 625 Kimbark | Des Moines | IA | 50311 | ....
 Início →
              Maumau | Maurício | 524 Rua dos Controles | Apucarana | PR | 86800 | .....
              Mantovani | Rafael | 14 Rua Cobol | Rolandópolis | PR | 86605 | .....
Meio
              Morrison | Sebatian | 9035 South Hill | Forest Village | OK | 74820 | .....
              Da Salsicha | Tamara | 10 Compila Street | Londrina PR | 85800 | .....
```

#### Achamos:))

```
/* Função para busca binária em arquivos
```

#### Parametros:

- arq: arquivo com os registros já ordenados
- chave: a chave do registro que procuramos
- reg: objeto para retornar o registro encontrado

#### Retorno (booleano):

- true: se encontrar o registro e salvar o conteúdo corretamente em reg
- false: se não encontrar o registro, e der algum erro durante o processo de leitura

bool buscaBinaria (FILE \*arq, char\* chave, Registro\* reg);

### Busca binária

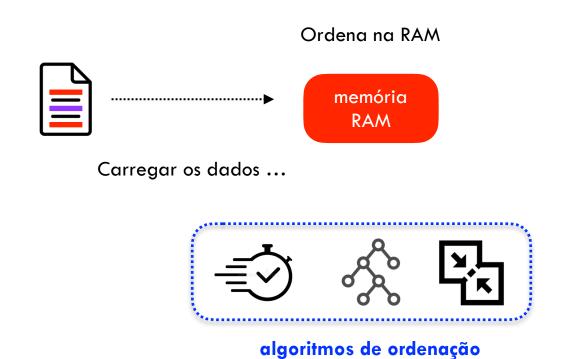
```
bool buscaBinaria (FILE *arq, char* chave, Registro* reg) {
 2.
          inicio = 0;
 3.
          fim = numeroRegistros(arq) - 1;
          enquanto (inicio <= fim) {</pre>
 4.
             meio = (inicio + fim)/2
 5.
             reg = lerRegistroComRRN(arq, meio)
 6.
             se reg.chave == chave return true;
 7.
 8.
             se reg.chave < chave: fim = meio -1;</pre>
             senão: inicio = meio + 1;
 9.
10.
          return false;
11.
12. }
```

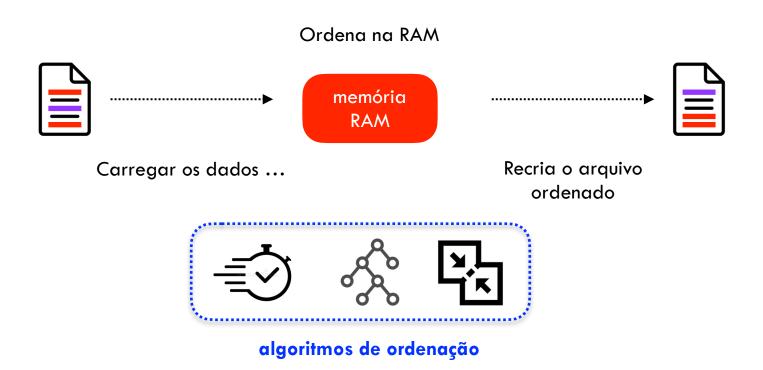
### Busca binária

- É atrativa:
  - mas só funciona se os registros estiveram ordenados
  - ordenar os registros com base em uma chave
  - ordenação é custosa, mas também um processo muito importante no processamento de arquivos









Limitações:

#### Limitações:

- busca binária ainda realiza muitas operações de leitura
- manter o arquivo sempre ordenado é custoso
- ordenação interna só funciona com arquivos pequenos

### Exercícios

1) Implemente uma função que realiza a busca binária em um arquivo de registros com tamanho fixo. Use os registros de heróis da aula passada como formato de registro (gravação/record).

```
/* Parâmetros:

arq: arquivo de entrada com os registros gravados

key: chave no formato canônico que será procurada

hero: registro retornado, quando encontrado

Returns: retorna true se encontrou

retorna false se não deletou, pa o registro não foi encontrado */

bool binarySearchFile (FILE *arq, char* key, Heroi* hero);
```

### Exercícios

### Roteiro

- 1 Introdução
- 2 Busca Binária
- 3 Key Sorting
- 4 Revisão
- 5 Referências

- KeySorting: ordenação por chaves
  - tentar solucionar o problema da ordenação interna
  - tag sort
  - ideia: quando queremos ordenar registros, ordenamos mesmo apenas as chaves

- KeySorting: ordenação por chaves
  - assumir que temos registros de tamanho fixo (pode funcionar para tamanhos variados também)
  - registro de cabeçalho no arquivo (header) com o número de registros do arquivo
  - algoritmo usa um vetor de pares ordenados {chave, RRN}
    - esse vetor se chama KEYNODES []

vetor KEYNODES []

#### Arquivo com Registros

#### Key

RAFAEL MANTOVANI
ANDREZÃO DAMASSA
TAMARA DA SALSICHA
FERNANDO VEGANO
MAURICIO MAUMAU
ERI NALDO

Criar chaves únicas

vetor KEYNODES [ ]

#### Arquivo com Registros

#### Key

RAFAEL MANTOVANI ANDREZÃO DAMASSA TAMARA DA SALSICHA FERNANDO VEGANO MAURICIO MAUMAU ERI NALDO

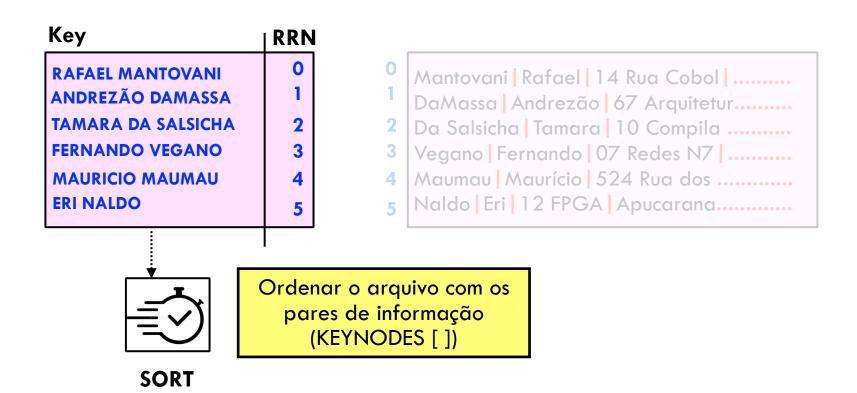
#### vetor KEYNODES []

#### Arquivo com Registros

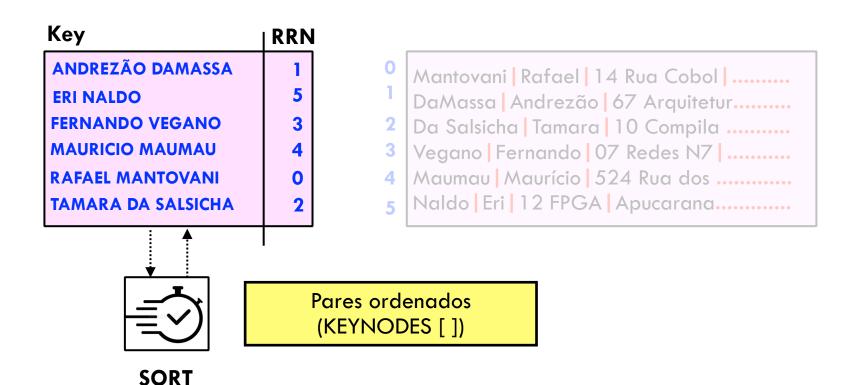
Key	RRN
RAFAEL MANTOVANI	0
ANDREZÃO DAMASSA	1
TAMARA DA SALSICHA	2
FERNANDO VEGANO	3
MAURICIO MAUMAU	4
ERI NALDO	5

Associar com RRN

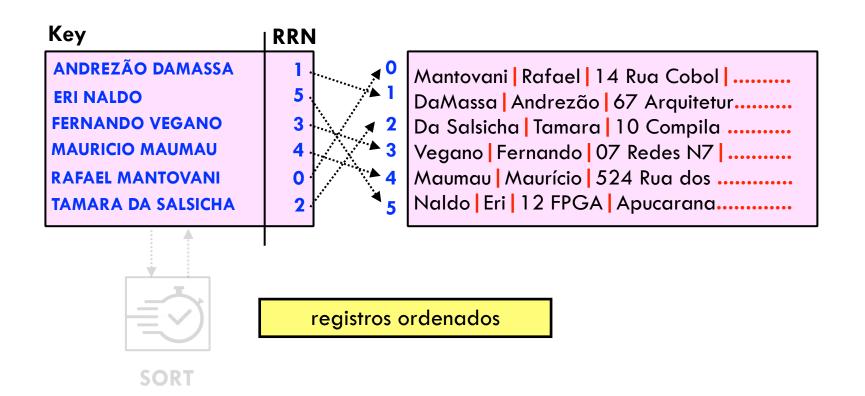
#### vetor KEYNODES []



#### vetor KEYNODES []

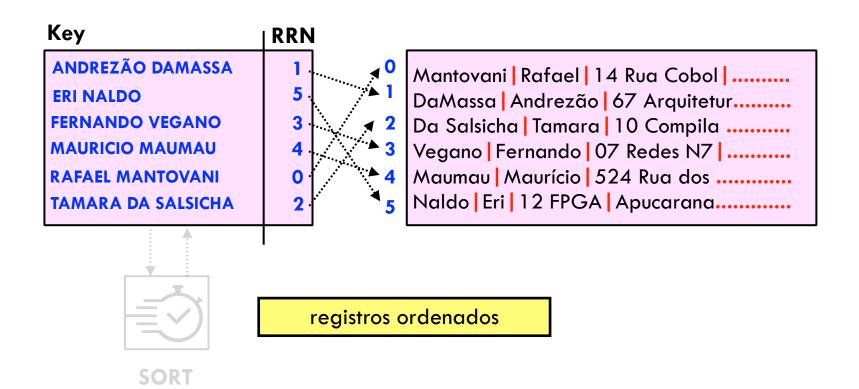


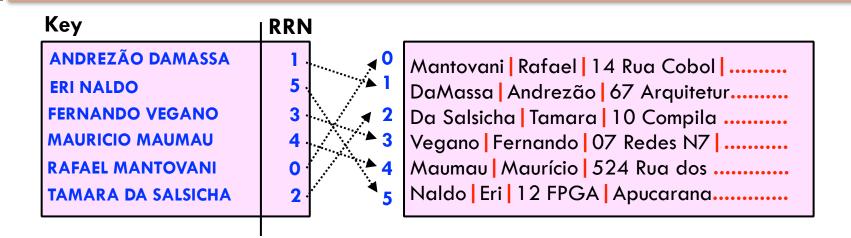
#### vetor KEYNODES []



#### vetor KEYNODES [] (RAM)

#### Arquivo com Registros (Disco)





Gerar um novo arquivo (ordenado)



Deleta os arquivos anteriores



Procura por registros

Arquivo ordenado

```
/* Função para ordenação via KeySorting
```

#### Parametros:

- arq: arquivo com os registros a serem lidos e ordenados
- arqSaida: nome do novo arquivo que será gerado com os registros ordenados

```
void keySort (FILE *arq, char* arqSaida);
```

```
void keySort (FILE *arg, char* argSaida) {
         N = numeroRegistros(arq)
 2.
 3.
        // cada KEYNODE é uma struct com = {chave, RRN}
 4.
        criar o vetor KEYNODES [N]
         para cada registro no arq:
 5.
             reg = lerRegistroComRRN(arq, i)
             KEYNODES[i] = \{\text{reg.chave, i}\} // i = RRN
7.
 8.
         ordenar KEYNODES[] com base nas chaves
 9.
         criar arquivo de saída arqSaida
         acessar o vetor ordenado KEYNODES[]
10.
         para cada posição de KEYNODES[]
11.
             reg = IerRegistroComRRN(arq, KEYNODES[i]) // RRN já ordenado
12.
13.
             gravar o registro no novo arquivo
14. }
```

Limitações:

#### Limitações:

- ordernar é custoso
- durante a ordenação, lemos os registros duas vezes do arquivo de entrada (nunca desejado!)
- criar o arquivo ordenado requer muitas operações de busca, para ter os valores ordenados no arquivo de saída (não desejado)
- basicamente, fazemos tudo em dobro!

Como melhorar?

### Roteiro

- 1 Introdução
- 2 Busca Binária
- **3** Key Sorting
- 4 Revisão
- 5 Referências

### Revisão

Busca binária

Ordenação Interna

KeySorting

### Revisão

- Busca binária
  - requer o arquivo ordenado
  - ainda custosa, mas melhor que a busca sequencial
- Ordenação Interna
  - move os registros para a memória e os ordena
  - só funciona para arquivos pequenos
- KeySorting
  - resolve o problema da ordenação
  - mas gasta muitas operações para criar o arquivo já ordenado

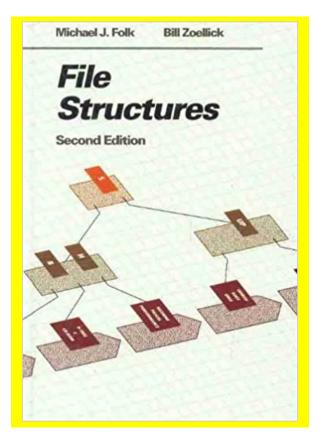
### Roteiro

- 1 Introdução
- 2 Busca Binária
- **3** Key Sorting
- 4 Revisão
- 5 Referências

## Referências sugeridas

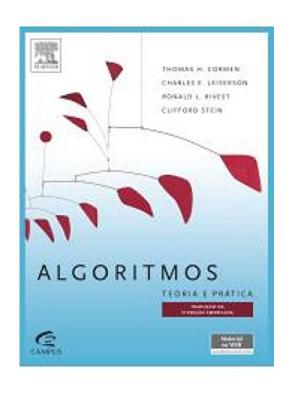


[Ziviani, 2010]



[Folk & Zoellick, 1992]

### Referências sugeridas



[Cormen et al, 2018]



[Drozdek, 2017]

# Perguntas?

Prof. Rafael G. Mantovani

rafaelmantovani@utfpr.edu.br