

EDCO4B

ESTRUTURAS DE DADOS 2

Aula 12 - Estruturas de Indexação
de Dados

Prof. Rafael G. Mantovani

Licença

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



maiores informações:

https://creativecommons.org/licenses/by-nc-nd/4.0/deed.pt_BR

Roteiro



- 1** Introdução
- 2** Estrutura de Índices
- 3** Operações
- 4** Revisão
- 5** Referências

Roteiro

- 1** Introdução
- 2** Estrutura de Índices
- 3** Operações
- 4** Revisão
- 5** Referências

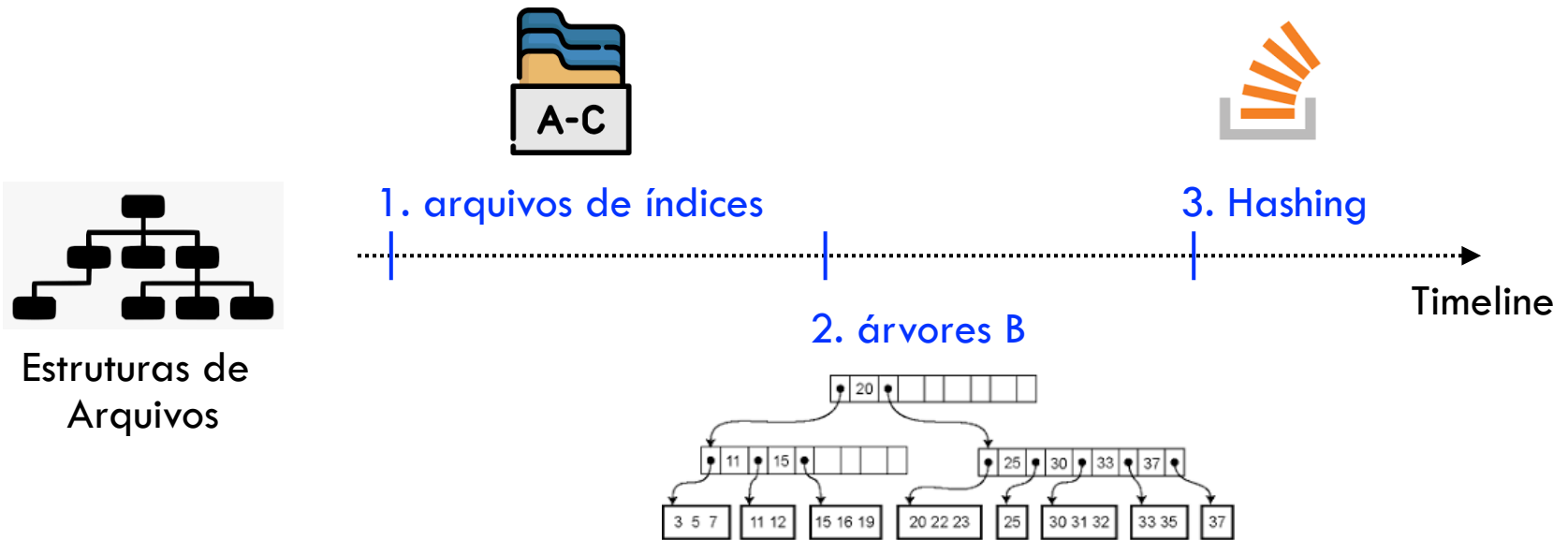
Introdução



- O que já fizemos:

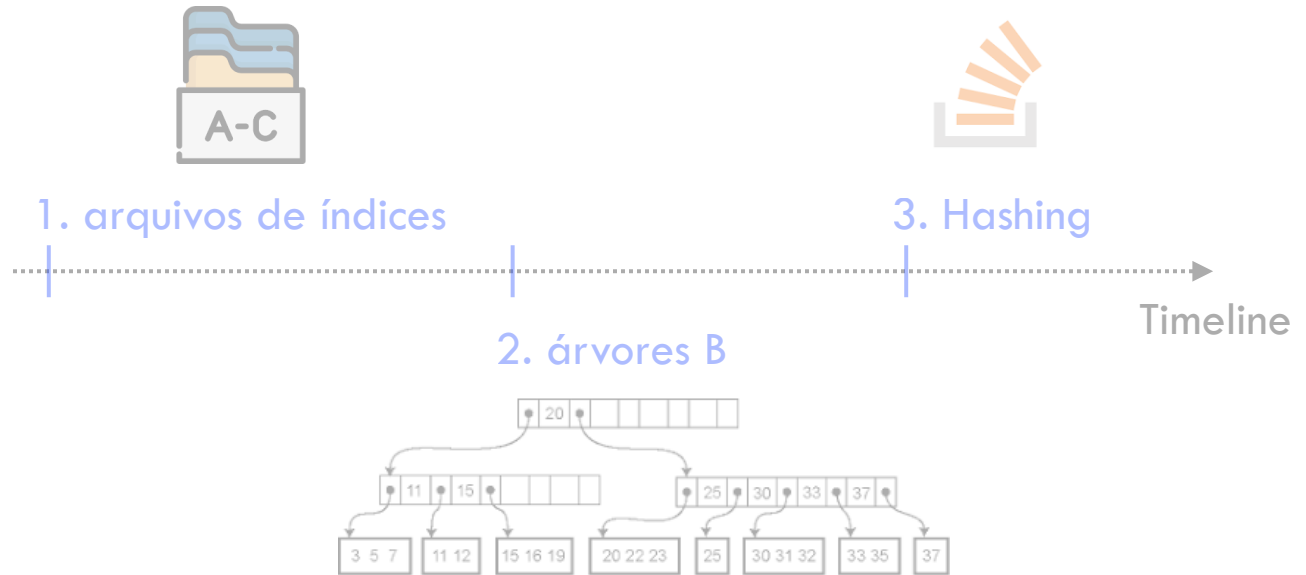
Introdução

- O que já fizemos:



Introdução

- O que já fizemos:

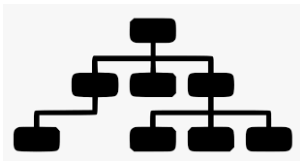


Introdução

- ○ que já 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
 - procurar informações / ordenar registros

Introdução

- O que já fizemos:



Estruturas de Arquivos

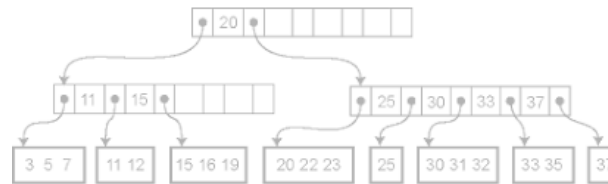


1. arquivos de índices



3. Hashing

2. árvores B



Timeline

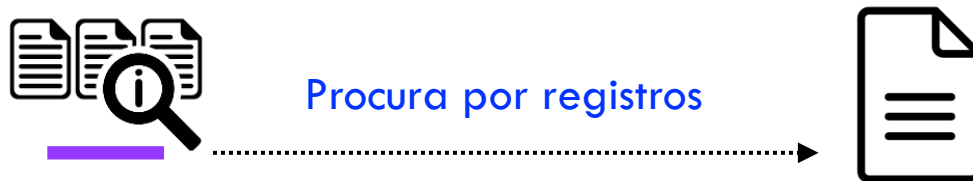
Roteiro



- 1 Introdução
- 2 Estrutura de Índices
- 3 Operações
- 4 Revisão
- 5 Referências

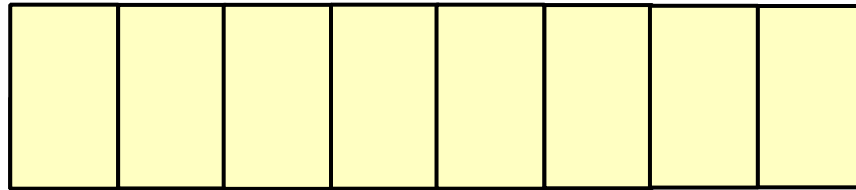
Índice

- **Índice:** estrutura de acesso auxiliar usada para melhorar o desempenho na recuperação de registros
- Pesquisa



- restringida a um subconjunto dos registros, em contrapartida à análise do conjunto completo
- realizada em resposta a certas condições

Índice



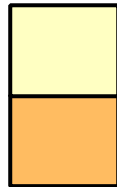
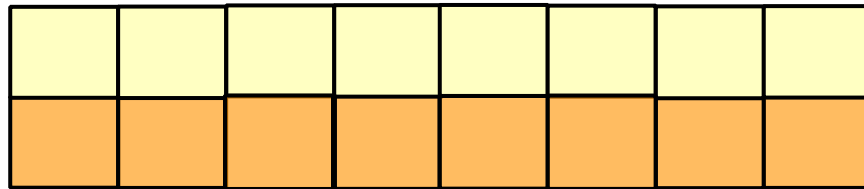
Índice



Índice



Índice



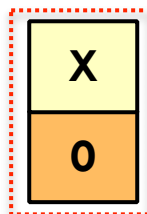
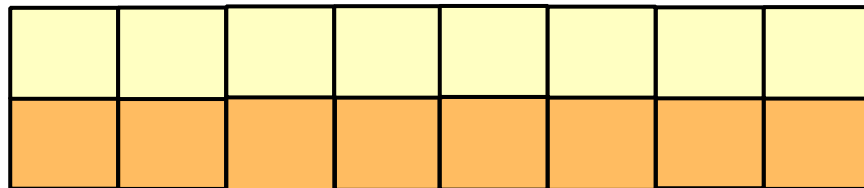
Índice





Chave de busca

Índice



Chave de busca

RRN ou *byte offset*

Índice



X	A	R	L	Y	Q	W	E
0	1	2	3	4	5	6	7



X
0

Chave de busca

RRN ou *byte offset*

Índice



X	A	R	L	Y	Q	W	E
0	1	2	3	4	5	6	7

Chave de busca

RRN ou *byte offset*

Índice



A	E	L	Q	R	X	Y	W
1	7	3	5	2	0	4	6

Chave de busca

RRN ou *byte offset*

Índice



A	E	L	Q	R	X	Y	W
1	7	3	5	2	0	4	6

Chave de busca

RRN ou *byte offset*

chaves ORDENADAS

Índice



A	E	L	Q	R	X	Y	W
1	7	3	5	2	0	4	6

Chave de busca

RRN ou *byte offset*

chaves ORDENADAS

Apesar de simples, índices proporcionam ferramentas poderosas para recuperar de registros

Índice



- **Exemplo:** registros de músicas

Índice

□ **Exemplo:** registros de músicas

LON	2312	Romeo and Juliet	Prokofiev ...
RCA	2626	Quartet in C sharp Minor ...	
WAR	23699	Touchstone	Corea ...
ANG	3795	Symphony No. 9	Beethoven ...
COL	38358	Nebraska	Springsteen ...
DG	18807	Symphony No. 9	Beethoven ...
MER	75016	Co1 d'or Suite	Rimsky ...
COL	31809	Symphony No. 9	Dvorak ...
DG	139201	Violin Concerto	Beethoven ...
FF	245	Good News	Sweet Honey in The ...

Índice

□ **Exemplo:** registros de músicas

LON	2312	Romeo and Juliet	Prokofiev ...
RCA	2626	Quartet in C sharp Minor ...	
WAR	23699	Touchstone	Corea ...
ANG	3795	Symphony No. 9	Beethoven ...
COL	38358	Nebraska	Springsteen ...
DG	18807	Symphony No. 9	Beethoven ...
MER	75016	Co1 d'or Suite	Rimsky ...
COL	31809	Symphony No. 9	Dvorak ...
DG	139201	Violin Concerto	Beethoven ...
FF	245	Good News	Sweet Honey in The ...

Arquivo de dados (disco)

Índice

□ **Exemplo:** registros de músicas

32	LON 2312 Romeo and Juliet Prokofiev ...
77	RCA 2626 Quartet in C sharp Minor ...
132	WAR 23699 Touchstone Corea ...
167	ANG 3795 Symphony No. 9 Beethoven ...
211	COL 38358 Nebraska Springsteen ...
256	DG 18807 Symphony No. 9 Beethoven ...
300	MER 75016 Co1 d'or Suite Rimsky ...
353	COL 31809 Symphony No. 9 Dvorak ...
396	DG 139201 Violin Concerto Beethoven ...
442	FF 245 Good News Sweet Honey in The ...

Arquivo de dados (disco)

Índice

□ Exemplo: registros de músicas

32	LON 2312 Romeo and Juliet Prokofiev ...
77	RCA 2626 Quartet in C sharp Minor ...
132	WAR 23699 Touchstone Corea ...
167	ANG 3795 Symphony No. 9 Beethoven ...
211	COL 38358 Nebraska Springsteen ...
256	DG 18807 Symphony No. 9 Beethoven ...
300	MER 75016 Co1 d'or Suite Rimsky ...
353	COL 31809 Symphony No. 9 Dvorak ...
396	DG 139201 Violin Concerto Beethoven ...
442	FF 245 Good News Sweet Honey in The ...

chaves **Arquivo de dados (disco)**

Índice

Exemplo: registros de músicas

Arquivo de índice
(disco)

32	LON 2312 Romeo and Juliet Prokofiev ...
77	RCA 2626 Quartet in C sharp Minor ...
132	WAR 23699 Touchstone Corea ...
167	ANG 3795 Symphony No. 9 Beethoven ...
211	COL 38358 Nebraska Springsteen ...
256	DG 18807 Symphony No. 9 Beethoven ...
300	MER 75016 Co1 d'or Suite Rimsky ...
353	COL 31809 Symphony No. 9 Dvorak ...
396	DG 139201 Violin Concerto Beethoven ...
442	FF 245 Good News Sweet Honey in The ...

chaves

Arquivo de dados (disco)

Índice

□ **Exemplo:** registros de músicas

LON2312
RCA2626
WAR23699
ANG3795
COL38358
DG18807
MER75016
COL31809
DG139201
FF245

**Arquivo de índice
(disco)**

32	LON 2312 Romeo and Juliet Prokofiev ...
77	RCA 2626 Quartet in C sharp Minor ...
132	WAR 23699 Touchstone Corea ...
167	ANG 3795 Symphony No. 9 Beethoven ...
211	COL 38358 Nebraska Springsteen ...
256	DG 18807 Symphony No. 9 Beethoven ...
300	MER 75016 Co1 d'or Suite Rimsky ...
353	COL 31809 Symphony No. 9 Dvorak ...
396	DG 139201 Violin Concerto Beethoven ...
442	FF 245 Good News Sweet Honey in The ...

Arquivo de dados (disco)

Índice

□ **Exemplo:** registros de músicas

LON2312	32
RCA2626	77
WAR23699	132
ANG3795	167
COL38358	211
DG18807	256
MER75016	300
COL31809	353
DG139201	396
FF245	442

**Arquivo de índice
(disco)**

32	LON 2312 Romeo and Juliet Prokofiev ...
77	RCA 2626 Quartet in C sharp Minor ...
132	WAR 23699 Touchstone Corea ...
167	ANG 3795 Symphony No. 9 Beethoven ...
211	COL 38358 Nebraska Springsteen ...
256	DG 18807 Symphony No. 9 Beethoven ...
300	MER 75016 Co1 d'or Suite Rimsky ...
353	COL 31809 Symphony No. 9 Dvorak ...
396	DG 139201 Violin Concerto Beethoven ...
442	FF 245 Good News Sweet Honey in The ...

Arquivo de dados (disco)

Índice

□ **Exemplo:** registros de músicas

LON2312	32
RCA2626	77
WAR23699	132
ANG3795	167
COL38358	211
DG18807	256
MER75016	300
COL31809	353
DG139201	396
FF245	442

**Arquivo de índice
(disco)**

32	LON 2312 Romeo and Juliet Prokofiev ...
77	RCA 2626 Quartet in C sharp Minor ...
132	WAR 23699 Touchstone Corea ...
167	ANG 3795 Symphony No. 9 Beethoven ...
211	COL 38358 Nebraska Springsteen ...
256	DG 18807 Symphony No. 9 Beethoven ...
300	MER 75016 Co1 d'or Suite Rimsky ...
353	COL 31809 Symphony No. 9 Dvorak ...
396	DG 139201 Violin Concerto Beethoven ...
442	FF 245 Good News Sweet Honey in The ...

Arquivo de dados (disco)

Índice

□ **Exemplo:** registros de músicas

ANG3795	167
COL31809	353
COL38358	211
DG139201	396
DG18807	256
FF245	442
LON2312	32
MER75016	300
RCA2626	77
WAR23699	132

**Arquivo de índice
(disco)**

32	LON 2312 Romeo and Juliet Prokofiev ...
77	RCA 2626 Quartet in C sharp Minor ...
132	WAR 23699 Touchstone Corea ...
167	ANG 3795 Symphony No. 9 Beethoven ...
211	COL 38358 Nebraska Springsteen ...
256	DG 18807 Symphony No. 9 Beethoven ...
300	MER 75016 Co1 d'or Suite Rimsky ...
353	COL 31809 Symphony No. 9 Dvorak ...
396	DG 139201 Violin Concerto Beethoven ...
442	FF 245 Good News Sweet Honey in The ...

Arquivo de dados (disco)

Índice

valores ordenados

ANG3795	167
COL31809	353
COL38358	211
DG139201	396
DG18807	256
FF245	442
LON2312	32
MER75016	300
RCA2626	77
WAR23699	132

Archivo de índice
(disco)

32	LON 2312 Romeo and Juliet Prokofiev ...
77	RCA 2626 Quartet in C sharp Minor ...
132	WAR 23699 Touchstone Corea ...
167	ANG 3795 Symphony No. 9 Beethoven ...
211	COL 38358 Nebraska Springsteen ...
256	DG 18807 Symphony No. 9 Beethoven ...
300	MER 75016 Co1 d'or Suite Rimsky ...
353	COL 31809 Symphony No. 9 Dvorak ...
396	DG 139201 Violin Concerto Beethoven ...
442	FF 245 Good News Sweet Honey in The ...

Archivo de datos (disco)

Índice

valores ordenados

ANG3795	167
COL31809	353
COL38358	211
DG139201	396
DG18807	256
FF245	442
LON2312	32
MER75016	300
RCA2626	77
WAR23699	132

Archivo de índice
(disco)

registros desordenados

32	LON 2312 Romeo and Juliet Prokofiev ...
77	RCA 2626 Quartet in C sharp Minor ...
132	WAR 23699 Touchstone Corea ...
167	ANG 3795 Symphony No. 9 Beethoven ...
211	COL 38358 Nebraska Springsteen ...
256	DG 18807 Symphony No. 9 Beethoven ...
300	MER 75016 Co1 d'or Suite Rimsky ...
353	COL 31809 Symphony No. 9 Dvorak ...
396	DG 139201 Violin Concerto Beethoven ...
442	FF 245 Good News Sweet Honey in The ...

Archivo de datos (disco)

Índice



- **Estrutura:**

Índice

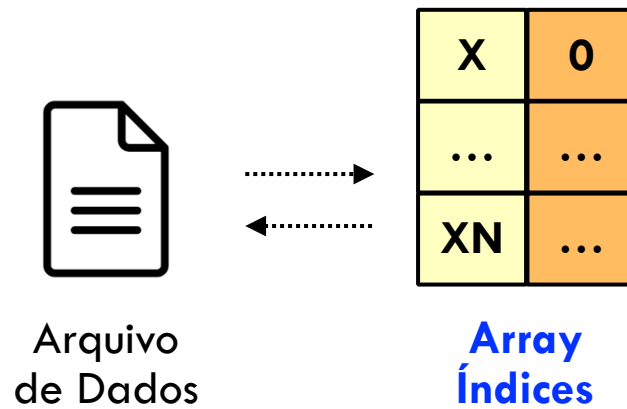
□ Estrutura:



**Arquivo
de Dados**

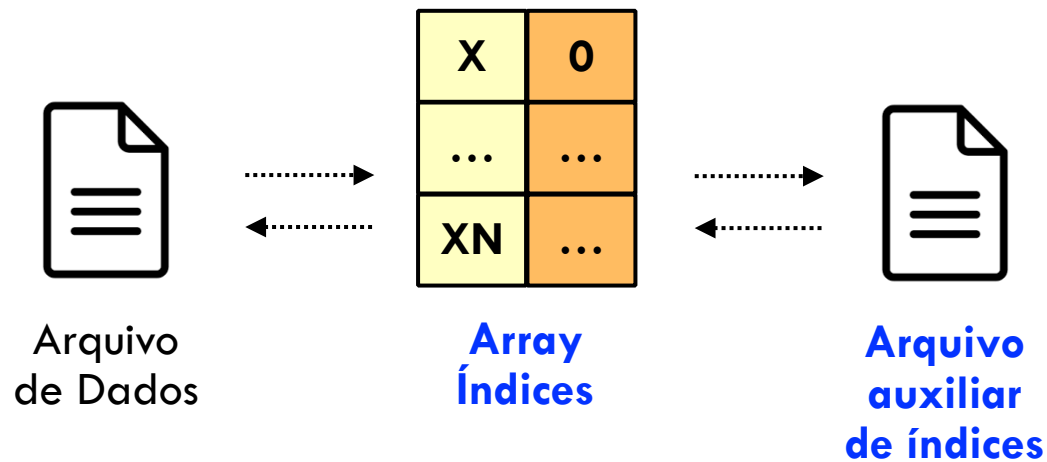
Índice

□ Estrutura:



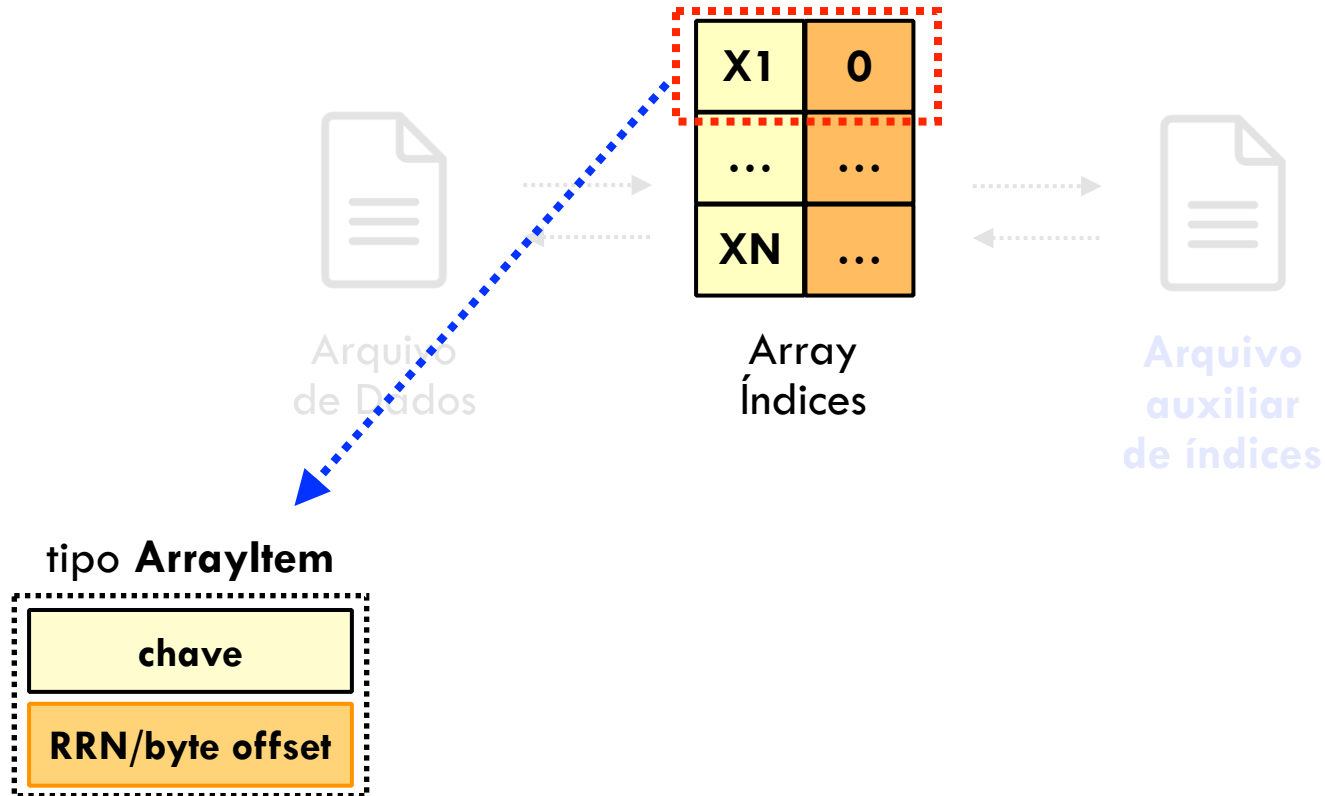
Índice

□ Estrutura:



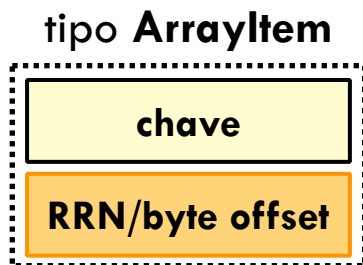
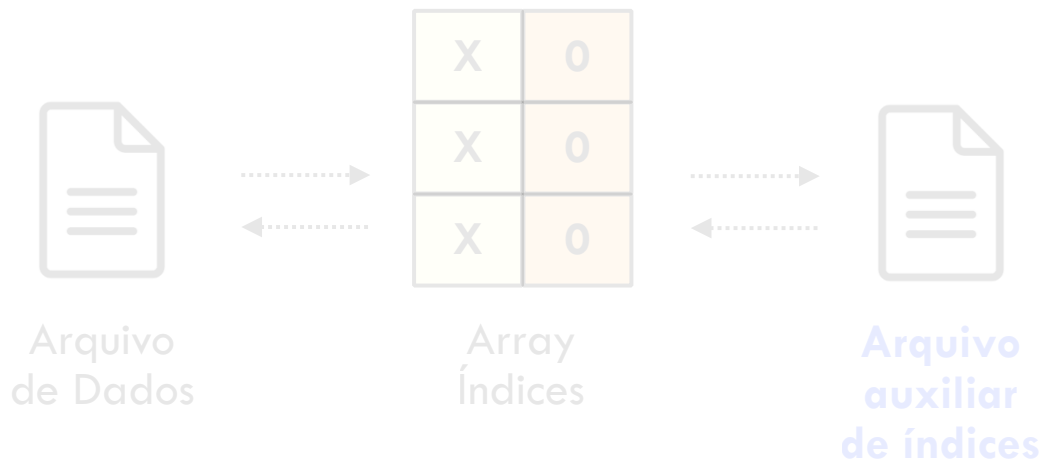
Índice

□ Estrutura:



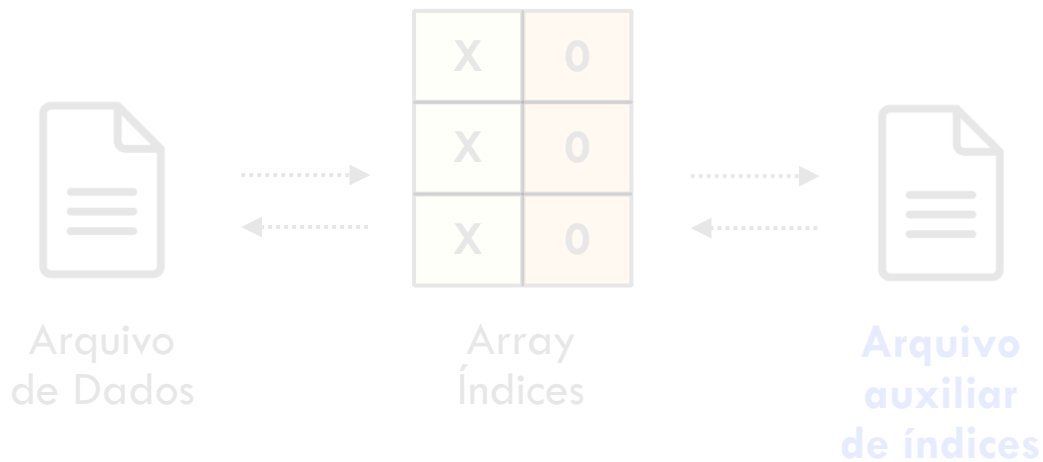
Índice

□ Estrutura:

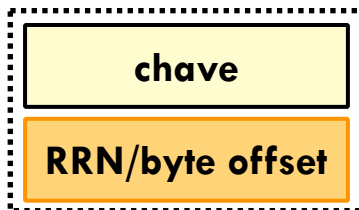


Índice

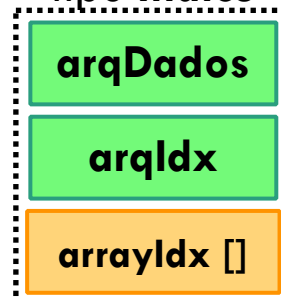
□ Estrutura:



tipo **ArrayItem**



tipo **Indice**



arquivo com os registros (disco)

arquivo auxiliar com os índices (disco)

índices ordenados (RAM)

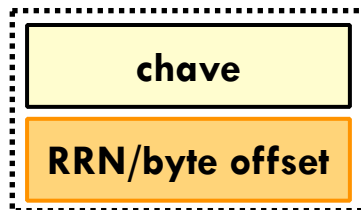
Índice

□ Estrutura:

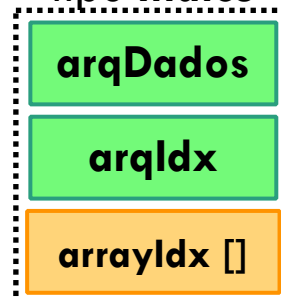
Dois tipos necessários: item para array (**ArrayItem**) e o índice (**Índice**) propriamente dito. Índice contém dois arquivos: um para os dados (registros), e outro auxiliar para gravar os índices. Além disso, mantém uma cópia do array de índices atualizada na memória (RAM) enquanto o programa executar.

de índices

tipo **ArrayItem**



tipo **Índice**



arquivo com os registros (disco)

arquivo auxiliar com os índices (disco)

índices ordenados (RAM)

Roteiro



- 1 Introdução
- 2 Estrutura de Índices
- 3 Operações
- 4 Revisão
- 5 Referências

Operações

Criar (S)

Criar (S, arq)

Carregar (S)

Inserir (S, reg)

Remover (S, k)

Pesquisar (S, k)

AtualizarIdx (S)

AtualizarReg (S, reg)

Destruir (S)

Operações

Criar (S)

Cria a estrutura vazia

Criar (S, arq)

Cria a estrutura de um arquivo com dados já existentes

Carregar (S)

Carrega o arquivo de índice na memória antes de usá-lo

Inserir (S, reg)

Insere um registro no arquivo (e Índice)

Remover (S, k)

Remove um registro do arquivo (e Índice)

Pesquisar (S, k)

Procura um registro no arquivo (e Índice)

AtualizarIdx (S)

Atualiza o Índice da memória para refletir as mudanças no arquivo

AtualizarReg (S, reg)

Atualiza um registro (e Índice)

Destruir (S)

Destrói a estrutura, grava o índice no arquivo auxiliar

Operações de modificação

Operações

lerRegistroComRRN (arq, RRN)

lerRegistroByteOffset (arq, byteOffset)

numeroRegistros (arq)

ordenaArrayIdx (arrayIdx[])

gravaIdxArqAux(arrayIdx[], arqIdx)

...

Operações

lerRegistroComRRN (arq, RRN)

lerRegistroByteOffset (arq, byteOffset)

numeroRegistros (arq)

ordenaArrayIdx (arrayIdx[])

gravaIdxArqAux(arrayIdx[], arqIdx)

...

Operações de auxiliares

Operações

lerRegistroComRRN (arq, RRN)

ler registro no arquivo baseado no RRN

lerRegistroByteOffset (arq, byteOffset)

ler registro no arquivo baseado no byte offset

numeroRegistros (arq)

ler número de registros do cabeçalho do arquivo (header)

ordenaArrayIdx (arrayIdx[])

Ordena o Índice na memória

gravalIdxArqAux(arrayIdx[], arqIdx)

Salva o Índice da memória para o arquivo auxiliar (disco)

...

Operações de auxiliares

Operações

Criar (S)

Criar (S, arq)

Carregar (S)

Inserir (S, reg)

Remover (S, k)

Pesquisar (S, k)

AtualizarIdx (S)

AtualizarReg (S, reg)

Destruir (S)

Operações de modificação

Operações de auxiliares

lerRegistroComRRN (arq, RRN)

lerRegistroByteOffset (arq, byteOffset)

numeroRegistros (arq)

ordenaArrayIdx (arrayIdx[])

gravalIdxArqAux(arrayIdx[], arqlIdx)

...

Exercícios

1) Implemente os tipos abstratos de dados necessários para definir uma estrutura de Índice. Defina também as assinaturas das funções para implementar cada uma das operações.

Pesquisa

Pesquisar (S , k)

: baseada na chave de busca (k)

Pesquisa

Pesquisar (S, k)

: baseada na chave de busca (k)

*/** Ideia principal do algoritmo */*

- * Encontra a posição da chave no **Índice**
- * Obtém o RRN ou byte offset do registro correspondente à posição encontrada
- * Encontra o registro no **arquivo de dados**
- * Recupera o registro solicitado do arquivo de dados

Criar

Criar (S)

: cria Índice juntamente do arquivo de dados
(só o cabeçalho é gerado)

Criar (S, arq)

: cria Índice baseado em um arquivo já existente

Criar

Criar (S)

: cria Índice juntamente do arquivo de dados
(só o cabeçalho é gerado)

Criar (S, arq)

: cria Índice baseado em um arquivo já existente

*/** Ideia principal do algoritmo **/*

- * adiciona o registro de cabeçalho (tamanho de registro, reuso, quantidade, etc)
- * varrer sequencialmente o **arquivo de dados**
 - * gerar chave de busca de cada registro
 - * adicionar a chave + campo de referência no Índice

Inserção

Inserir (S, reg)

: inserir um registro no arquivo de dados. Modifica também o Índice

Inserção

Inserir (S, reg)

: inserir um registro no arquivo de dados. Modifica também o Índice



**Inserção de um novo
registro no arquivo
de dados**

1:1



X	0
..	..



**Inserção de um novo
registro no arquivo
de Índice**

Inserção

Inserir (S, reg)

: inserir um registro no arquivo de dados. Modifica também o Índice



Inserção de um novo
registro no arquivo
de dados

1:1



X	0
..	..



Inserção de um novo
registro no arquivo
de Índice

Arquivo não ordenado: no final
do arquivo ou com **reuso**

Inserção

Inserir (S, reg)

: inserir um registro no arquivo de dados. Modifica também o Índice



Inserção de um novo
registro no arquivo
de dados

Arquivo não ordenado: no final
do arquivo ou com **reuso**

1:1



X	O
..	..



Inserção de um novo
registro no arquivo
de Índice

Necessidade de reorganização
do Índice, devido à **ordenação**
da chave

Inserção

Inserir (S, reg)

: inserir um registro no arquivo de dados. Modifica também o Índice

*/** Ideia principal do algoritmo */*

Remoção

Remover (S, k)

: remove registros no Índice devido às remoções no arquivo de dados

Remoção

Remover (S, k)

: remove registros no Índice devido às remoções no arquivo de dados



Remoção de um
registro no arquivo
de dados

1:1



X	0
..	..



Remoção de um
registro no arquivo
de Índice

Remoção

Remover (S, k)

: remove registros no Índice devido às remoções no arquivo de dados



Remoção de um
registro no arquivo
de dados

1:1



X	0
..	..



Remoção de um
registro no arquivo
de Índice

* Lógica: **reaproveitamento** de
espaços

Remoção

Remover (S, k)

: remove registros no Índice devido às remoções no arquivo de dados



Remoção de um
registro no arquivo
de dados

1:1



X	0
..	..



Remoção de um
registro no arquivo
de Índice

* Lógica: **reaproveitamento** de
espaços

* Lógica: do Índice (**RAM**)
* Física: deslocamento dos registros
(**arquivo**)

Remoção

Remover (S, k)

: remove registros no Índice devido às remoções no arquivo de dados

*/** Ideia principal do algoritmo */*

Atualização/Update

AtualizarReg (S, reg)

: modifica registros no Índice devido às modificações no arquivo de dados

Atualização/Update

AtualizarReg (S, reg)

: modifica registros no Índice devido às modificações no arquivo de dados



1:1



Mais comum: remoção
seguida de inserção

X	0
..	..

Atualização/Update

AtualizarReg (S, reg)

: modifica registros no Índice devido às modificações no arquivo de dados



1:1



Mais comum: remoção
seguida de inserção

X	0
..	..

- * **mudou campo chave?** : remodelação do Índice
- * **mudou campo não chave?** : ajuste do campo de referência se o registro mudar fisicamente no arquivo de dados

Atualização/Update

AtualizarReg (S, reg)

: modifica registros no Índice devido às modificações no arquivo de dados

*/** Ideia principal do algoritmo **/*

Atualiza Índice

AtualizarIdx (S)

: atualiza o arquivo de índice em disco com base no Índice em memória (reescrita)

Atualiza Índice

AtualizarIdx (S)

: atualiza o arquivo de índice em disco com base no Índice em memória (reescrita)

X	0
..	..

Índice
(RAM)

.....→
**atualiza arquivo aux
(depois de operações)**



Arquivo
de Índice
(disco)

Atualiza Índice

AtualizarIdx (S)

: atualiza o arquivo de índice em disco com base no Índice em memória (reescrita)

X	0
..	..

Índice
(RAM)

.....→
**atualiza arquivo aux
(depois de operações)**



Arquivo
de Índice
(disco)

Informação adicional (header): status (true/false) se está atualizado
* Ajuda a lidar com inconsistência nos índices (problemas, etc ...)

Atualiza Índice

AtualizarIdx (S)

: atualiza o arquivo de índice em disco com base no Índice em memória (reescrita)

X	0
..	..

Índice
(RAM)

atualiza arquivo aux
(depois de operações)



Arquivo
de Índice
(disco)

Informação adicional (header): **status** (**true/false**) se está atualizado
* Ajuda a lidar com inconsistência nos índices (problemas, etc ...)

Exercícios


2) Implemente uma estrutura de Índice e suas operações. Use o arquivo de heróis como exemplo de registro, e assuma que o arquivo de dados seja organizado em registros de tamanho fixo, mas com campos variáveis.

Exercícios

```
// struct para representar um Heroi
typedef struct {
    char primeiroNome[16];    // primeiro nome
    char sobrenome[16];       // último nome do heroi
    char nomeHeroi[16];       // nome de heroi, alias
    char poder[16];           // poder
    char fraqueza[21];        // fraqueza
    char cidade[21];          // cidade onde vive
    char profissao[21];       // profissão
} Heroi;
```

Roteiro



- 1 Introdução
 - 2 Estrutura de Índices
 - 3 Operações
 - 4 Revisão
 - 5 Referências
- 

Revisão



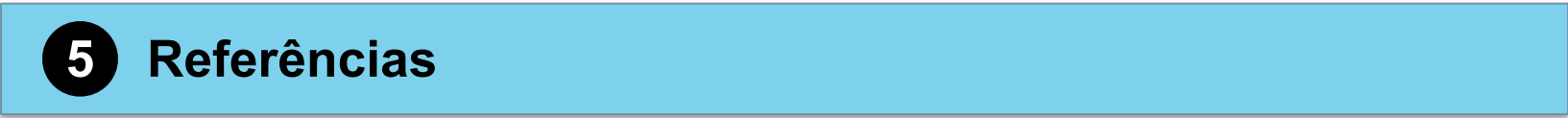
- **Índice:**

Revisão

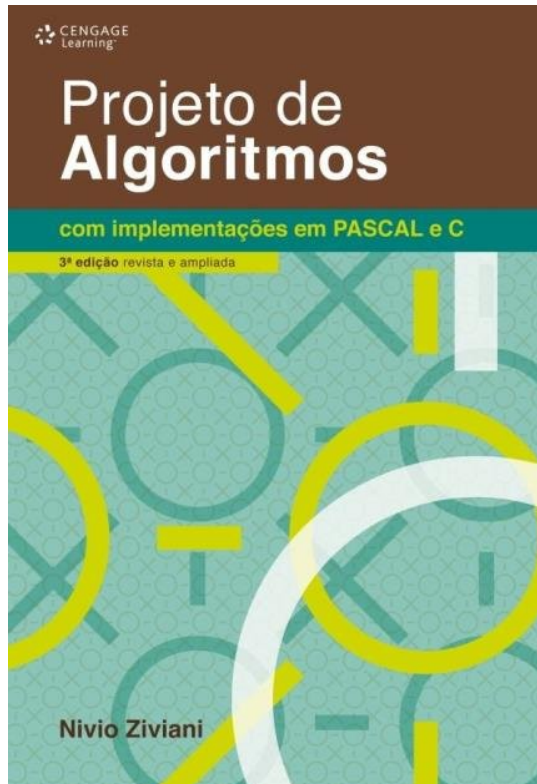
- **Índice:** estrutura de acesso auxiliar usada para melhorar o desempenho na recuperação de registros
 - permite buscar informações via busca binária usando “chaves”
 - se o índice cabe na memória, as operações podem ser feitas rapidamente com um arquivo auxiliar mantém o índice das informações

Roteiro

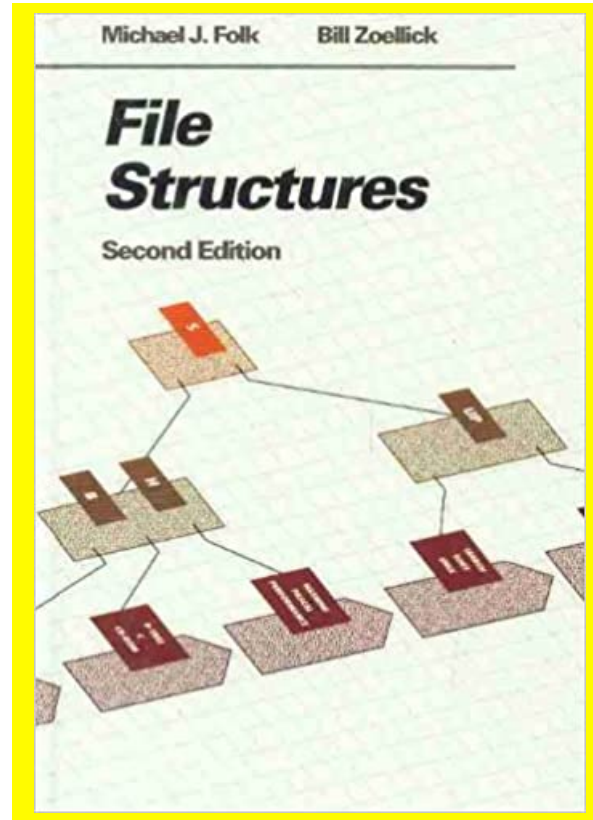


- 1** Introdução
 - 2** Estrutura de Índices
 - 3** Operações
 - 4** Revisão
 - 5** Referências
- 

Referências sugeridas

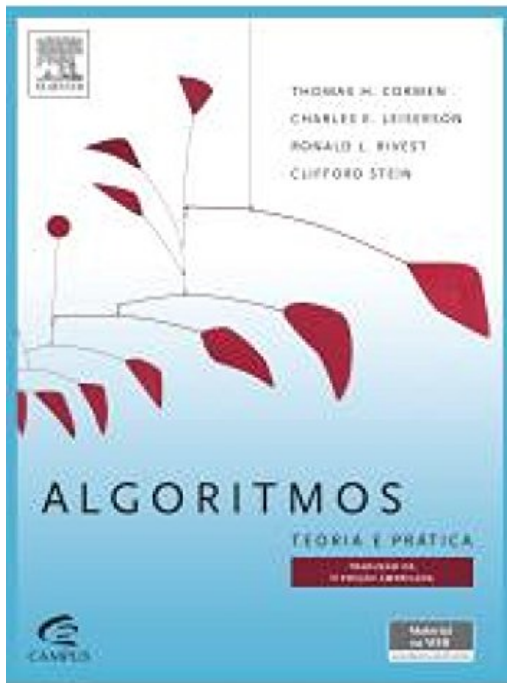


[Ziviani, 2010]



[Folk & Zoellick, 1992]

Referências sugeridas



[Cormen et al, 2018]



[Drozdek, 2017]

Perguntas?

Prof. Rafael G. **Mantovani**

rafaelmantovani@utfpr.edu.br