

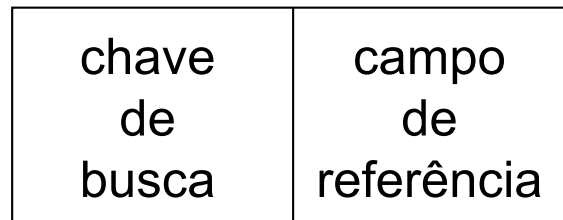
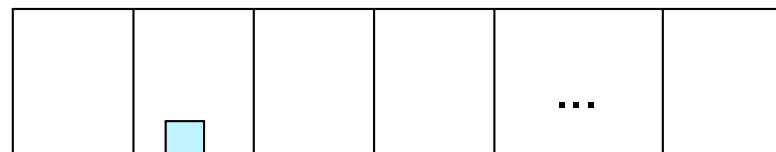
Estruturas de Indexação de Dados

Cristina Dutra de Aguiar Ciferri

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 Simples ou Linear



valores
ordenados

RRN (registros de tamanho fixo)

ou

byte offset (registros de tamanho variável)

apesar de simples,
índices proporcionam
ferramentas poderosas
para a recuperação
de registros

Índice Simples ou Linear

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

arquivo de índice

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 Coq 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 ...

chave 

arquivo de dados

arquivo auxiliar em disco

arquivo armazenado em disco

Índice Simples ou Linear

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

arquivo de índice

valores ordenados

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 Coq 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

geralmente registros desordenados

Índice Simples ou Linear

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

arquivo de índice

campos e

registros de tamanho fixo

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 Coq 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 ...

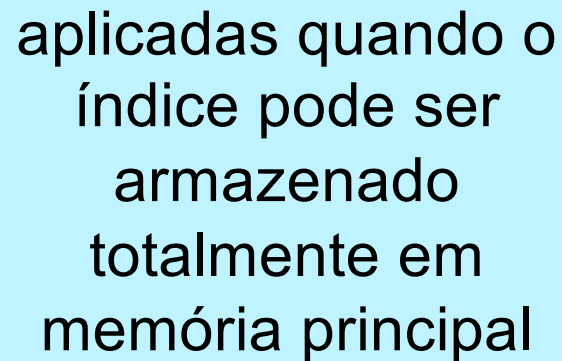
arquivo de dados

campos e

registros de tamanho fixo ou variável

Operações em um Índice

- Pesquisa
- Criação
- Inserção
- Remoção
- Atualização
- Destruição
- Carregamento
- Reescrita



aplicadas quando o índice pode ser armazenado totalmente em memória principal

Pesquisa

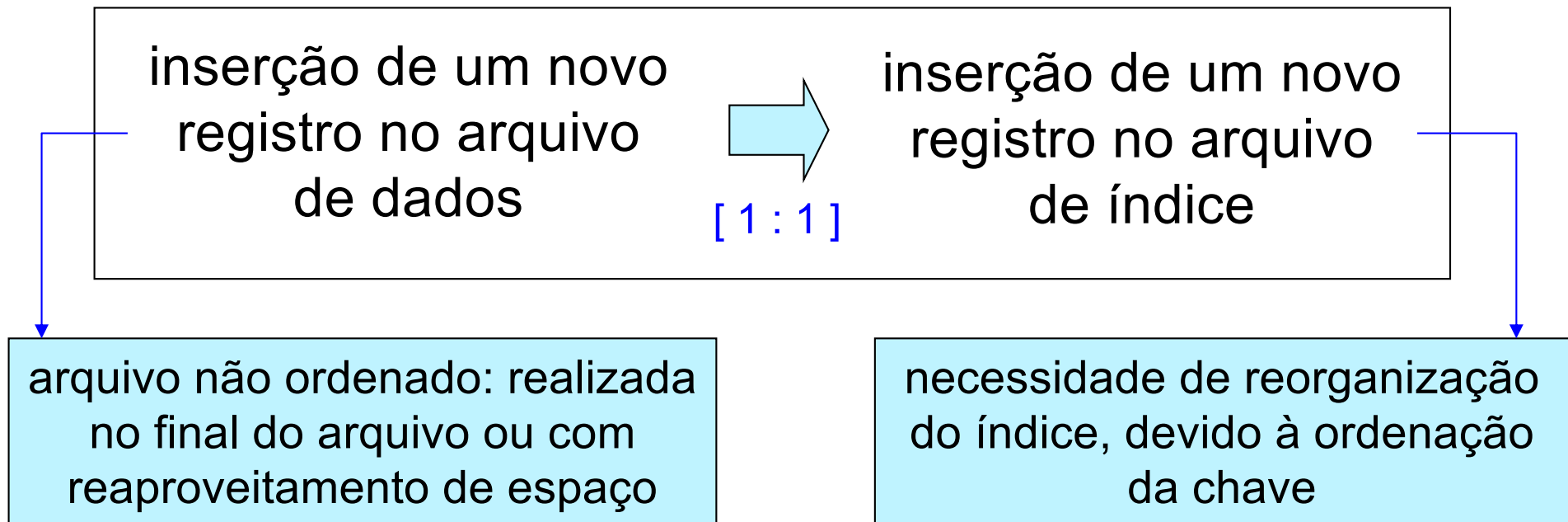
- Baseada na chave de busca
 - encontra a posição da chave no **arquivo de índice**
 - obtém o RRN ou o *byte offset* do registro correspondente à posição encontrada
 - encontra o registro no **arquivo de dados**
 - recupera o registro solicitado do arquivo de dados

Criação

- Cria o índice juntamente com a criação do arquivo de dados
 - apenas registro de cabeçalho
- Cria o índice baseado em um arquivo de dados já existente
 - registro de cabeçalho
 - demais registros (chave de busca + campo de referência), obtidos a partir de uma varredura no arquivo de dados

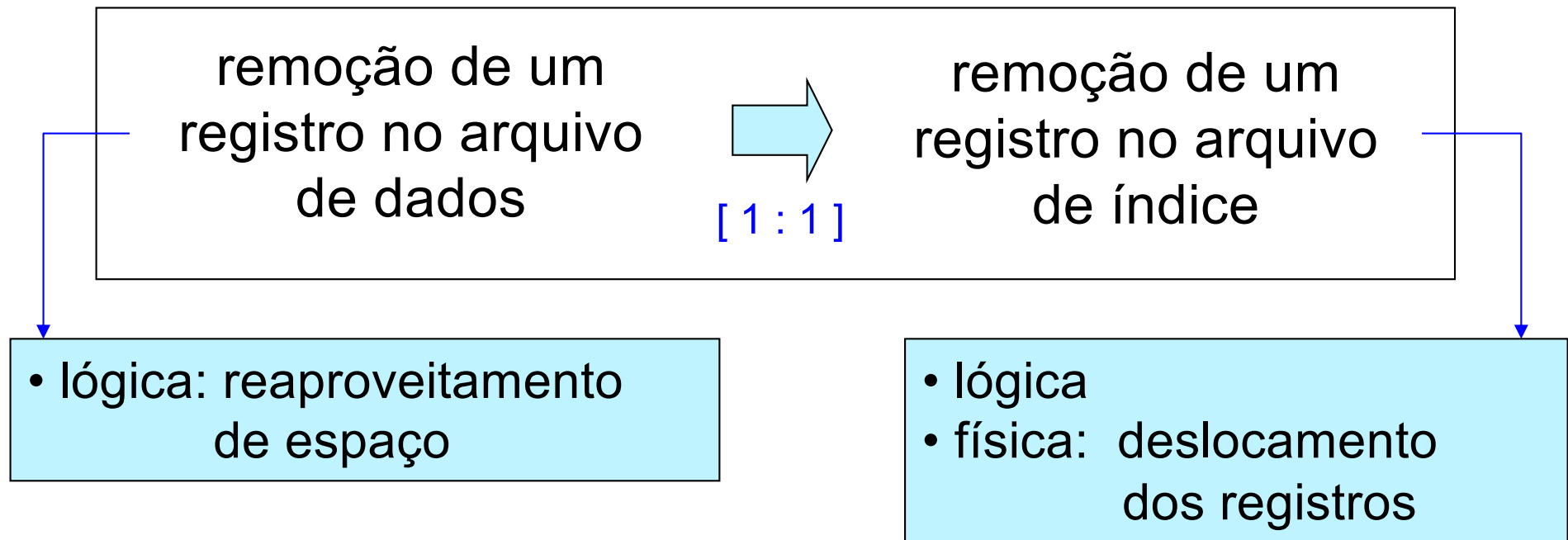
Inserção

- Adiciona registros no índice devido às inserções no arquivo de dados



Remoção

- Remove registros no índice devido às remoções no arquivo de dados



Atualização

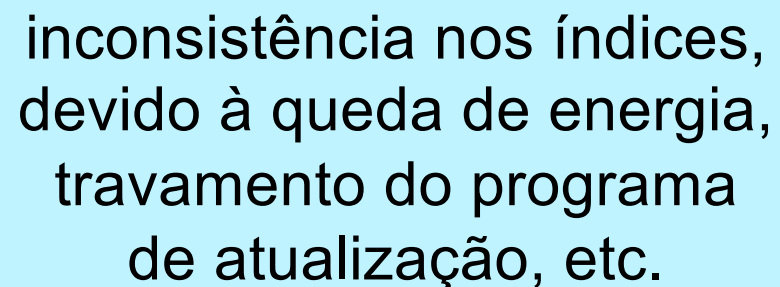
- Modifica registros no índice devido às modificações no arquivo de dados
- Tratamento
 - remoção seguida de inserção ← técnica mais utilizada
 - ou*
 - campo chave: reordenação do índice
 - campo não chave: ajuste do campo de referência se o registro mudar fisicamente no arquivo de dados

Carregamento

- Carrega o arquivo de índice na memória principal antes de usá-lo
- Passos
 - aponta para o primeiro registro do arquivo de índice em disco
 - varre o arquivo de índices sequencialmente
 - cria o índice em memória principal, em geral implementado como um vetor

Reescrita

- Atualiza o arquivo de índice em disco com base no arquivo de índice em memória principal, quando necessário
- Informação adicional
 - *status* no registro de cabeçalho
 - verdadeiro/falso



inconsistência nos índices,
devido à queda de energia,
travamento do programa
de atualização, etc.

Índice Simples ou Linear

- Primário
 - definido sobre um campo sem repetição
- Secundário
 - definido sobre um campo com repetição
 - tipos
 - fracamente ligado (*Loosely Binding*)
 - fortemente ligado (*Tight Binding*)

Índice Secundário Fracamente Ligado

Beethoven	ANG3795
Beethoven	DG139201
Beethoven	DG18807
Beethoven	RCA2626
Corea	WAR23699
Dvorak	COL31809
Prokofiev	LON2312
Rimsky	MER75016
Springsteen	COL38358
Sweet Honey In The	FF245

arquivo de índice secundário

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

arquivo de índice primário

32	LON 2312 Rom ...
77	RCA 2626 Quar ...
132	WAR 23699 Tou ...
167	ANG 3795 Sym ...
211	COL 38358 Nebr ...
256	DG 18807 Sym ...
300	MER 75016 Coq ...
353	COL 31809 Sym ...
396	DG 139201 Violin ...
442	FF 245 Good ...

arquivo de dados

Índice Secundário Fortemente Ligado

Beethoven	167
Beethoven	396
Beethoven	256
Beethoven	77
Corea	132
Dvorak	353
Prokofiev	32
Rimsky	300
Springsteen	211
Sweet Honey In The	442

arquivo de índice secundário

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 Coq 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

Repetição de Valores

- Problemas
 - **necessidade de armazenar** a mesma chave secundária várias vezes
 - **necessidade de reordenar** os índices sempre que um novo registro é inserido no arquivo
 - mesmo que esse registro tenha um valor de chave secundária já existente no arquivo
- Melhorias
 - solução 1. **vetores de tamanho fixo**
 - solução 2. **listas invertidas**

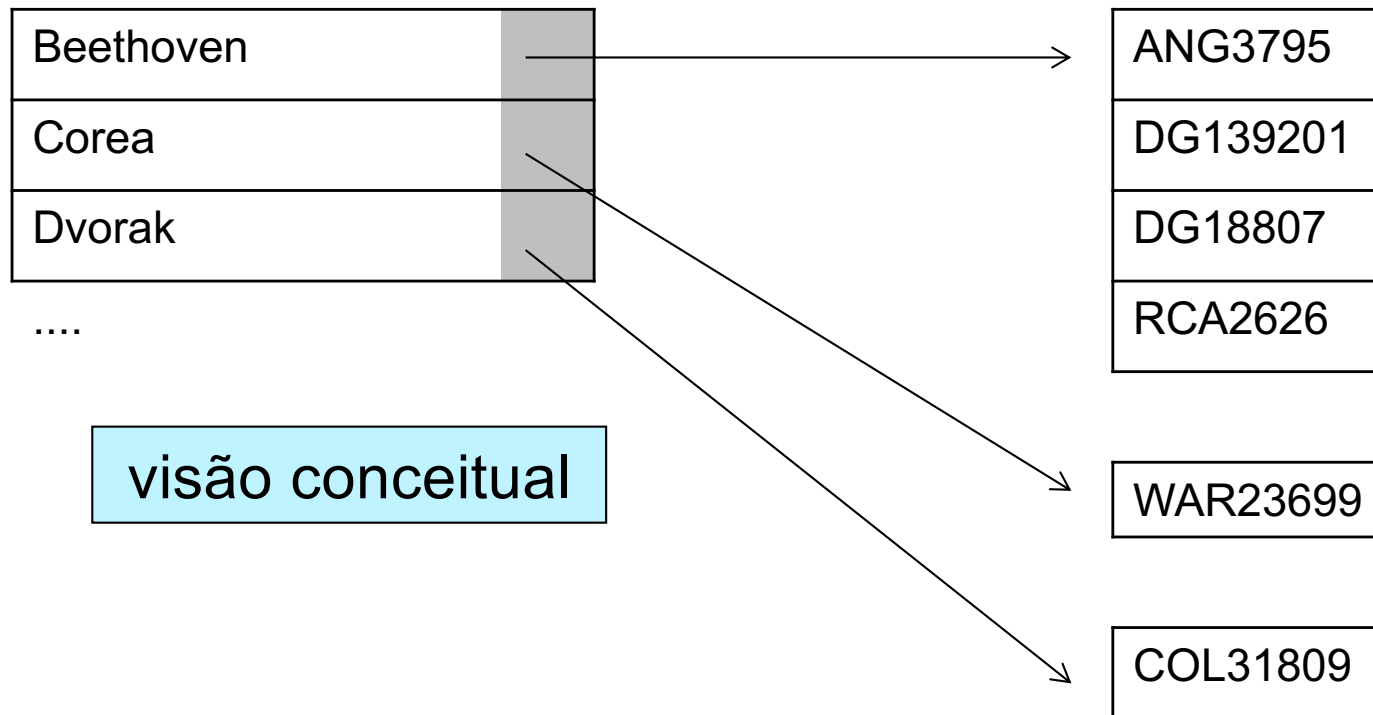
Vetores de Tamanho Fixo

- Associa um vetor de tamanho fixo a cada chave secundária

Beethoven	ANG3795	DG139201	DG18807	RCA2626
Corea	WAR23699			
Dvorak	COL31809			
Prokofiev	LON2312			
Rimsky	MER75016			
Springsteen	COL38358			
Sweet Honey In The	FF245			

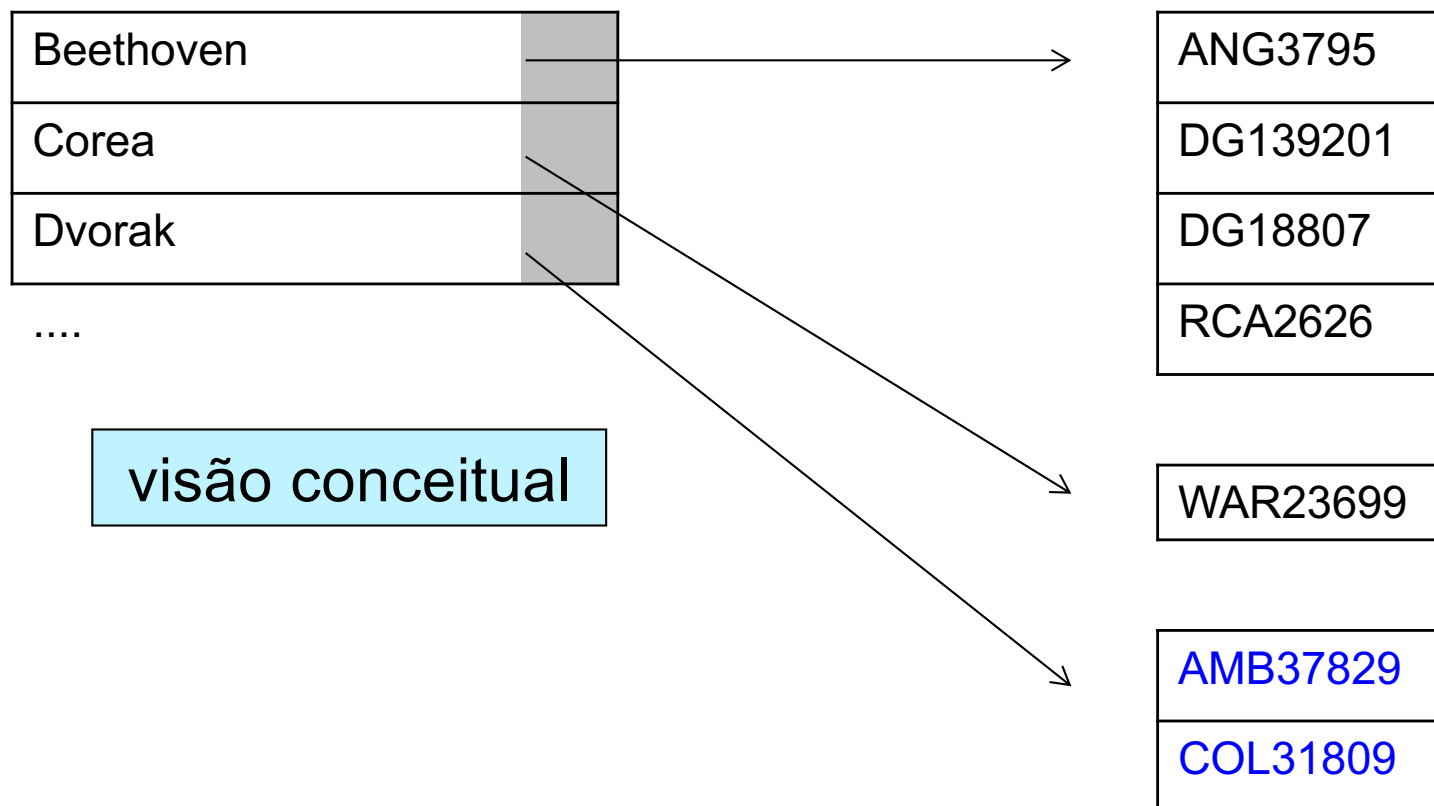
Listas Invertidas

- Associa uma lista encadeada das chaves primárias a cada chave secundária



Listas Invertidas

- Inserção de um novo registro: Dvorak



Índice Simples ou Linear

- Adequado quando cabe em **memória primária**
- Armazenamento em **memória secundária**
 - pode requerer vários acessos a disco, por causa da busca binária
 - pode ter manutenção cara, devido ao à adição e remoção de registros
 - requer o uso de outras organizações mais apropriadas, como **árvore-B**.