

Engenharia de Computação

Estrutura de Dados 2

Aula 12 – Indexação Secundária

Prof. Muriel de Souza Godoi
muriel@utfpr.edu.br

Índice Secundário

- Definido sobre uma **chave secundária**
 - podem ser definidos vários índices secundários para um mesmo arquivo de dados
- Tipos
 - **fracamente ligado** (*Loosely Binding*)
 - relaciona uma chave secundária à chave primária
 - **fortemente ligado** (*Tight Binding*)
 - relaciona uma chave secundária diretamente ao registro

Índice Secundário – Fracamente Ligado

índice secundário

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

índice primário

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

aquivo de dados

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 ...

Índice Secundário – Fracamente Ligado

Busca

- Passo 1
 - pesquisar o índice de chave secundária para encontrar a chave primária relacionada
- Passo 2
 - usar a chave primária para pesquisar o índice de chave primária para encontrar o byte offset (ou RRN) do registro no arquivo de dados
- Passo 3
 - recuperar o registro desejado

Índice Secundário – Fracamente Ligado

Inserção

- Passo 1
 - inserir o registro no arquivo de dados
- Passo 2
 - inserir a entrada correspondente no arquivo de índice primário
- Passo 3
 - inserir a entrada correspondente em cada arquivo de índice secundário
 - chaves duplicadas devem ser mantidas agrupadas e ordenadas

Índice Secundário – Fracamente Ligado

- **Remoção**

- **Estratégia 1:** Apagar **todas** as referências

- **Passo 1**

- remover o registro no **arquivo de dados**

- **Passo 2**

- remover a entrada correspondente no arquivo de **índice primário**

- **Passo 3**

- remover a entrada correspondente em cada arquivo de **índice secundário**

Índice Secundário – Fracamente Ligado

- **Remoção**
 - **Estratégia 2:** Apagar **algumas** referências
- **Passo 1**
 - remover o registro no **arquivo de dados**
- **Passo 2**
 - remover a entrada correspondente no arquivo de **índice primário**
- **Observações:**
 - Mantém a entrada correspondente no índice secundário
 - Pesquisa no índice primário acusará que o registro foi removido

Índice Secundário - Remoção

	apagar todas as referências	apagar algumas referências
vantagens	<ul style="list-style-type: none">• sem queda de desempenho na busca por registros removidos• índices permanecem do tamanho necessário	<ul style="list-style-type: none">• sem necessidade de reorganização a cada remoção• economia de tempo nas remoções
desvantagens	<ul style="list-style-type: none">• necessidade de reorganização a cada remoção• processo altamente custoso, devido à ordenação	<ul style="list-style-type: none">• com queda de desempenho na busca na busca por registros removidos• crescimento do tamanho dos índices e necessidade de reorganização periódica

Índice Secundário - Atualização

- Alteração do valor da **chave secundária**
 - reordenação do índice secundário
- Alteração do valor da **chave primária**
 - reordenação do índice primário
 - atualização dos índices secundários
 - reordenação dos índices secundários se houver repetição da chave secundária
- Alteração dos **demaís campos**
 - não afeta nenhum dos índices
 - mas, e se o **tamanho** do registro mudar?

Índice Secundário – Fortemente Ligado

índice secundário

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

aquivo de dados

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 ...

Índice Secundário – Fortemente Ligado

Busca

- Passo 1
 - pesquisar o **índice de chave secundária** para encontrar o *byte offset* (ou RRN) do registro no arquivo de dados
- Passo 2
 - recuperar o registro desejado

Índice Secundário – Fortemente Ligado

Inserção

- Passo 1
 - inserir o registro no **arquivo de dados**
- Passo 2
 - inserir a entrada correspondente em cada arquivo de **índice secundário**
 - chaves duplicadas devem ser mantidas agrupadas e ordenadas

Índice Secundário – Fortemente Ligado

Remoção

- Passo 1
 - remover o registro no arquivo de dados
- Passo 2
 - **Apagar todas as referências:** remover a entrada correspondente em cada arquivo de índice secundário
 - ou
 - **Apagar algumas referências:** manter a entrada correspondente em cada arquivo de índice secundário

Índice Secundário – Fortemente Ligado

Atualização

- Alteração do valor da **chave secundária**
 - reordenação do índice secundário
- Alteração dos **demaís campos**
 - não afeta nenhum dos índices
 - exceto, se o **tamanho** do registro mudar

Tipos de Índice Secundário

	Fracamente ligado	Fortemente ligado
vantagens	<ul style="list-style-type: none">• diminui custo de remoções na abordagem “<i>apagar algumas referencias</i>”• modificação no arquivo de dados afeta apenas o índice primário• menor complexidade de codificação	<ul style="list-style-type: none">• acesso direto → índice primário → arquivo de dados → índice secundário → arquivo de dados• melhor desempenho na busca
desvantagens	<ul style="list-style-type: none">• acesso indireto: → índice secundário → índice primário → arquivo de dados• queda do desempenho na busca	<ul style="list-style-type: none">• alto custo para modificações• modificação no arquivo de dados afeta todos os índices secundários• maior complexidade de codificação

Repetição de Chaves Secundárias

- 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

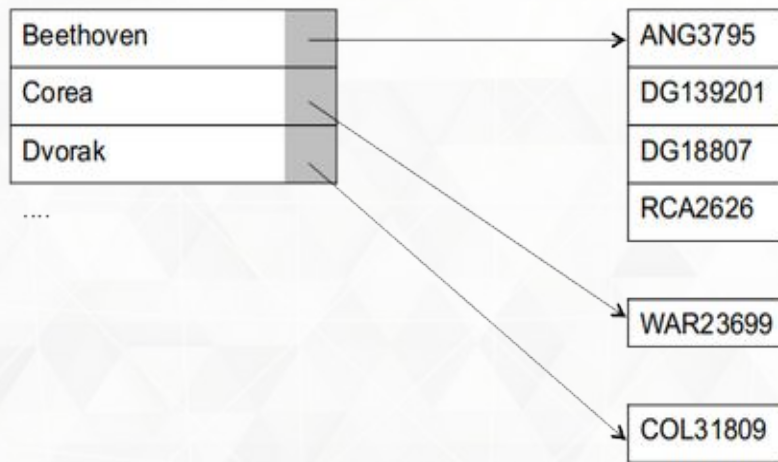
Beethoven	ANG3795	DG139201	DG18807	RCA2626
Corea	WAR23699			
Dvorak	COL31809			
Prokofiev	LON2312			
Rimsky	MER75016			
Springsteen	COL38358			
Sweet Honey In The	FF245			

Vetores de Tamanho Fixo

- Vantagem
 - não é necessário reordenar o índice secundário a cada inserção de chave secundária repetida
- Desvantagens
 - limitado a um número fixo de repetições
 - grande ocorrência de fragmentação interna no índice
 - pode não compensar a eliminação da duplicação de chaves

Listas Invertidas

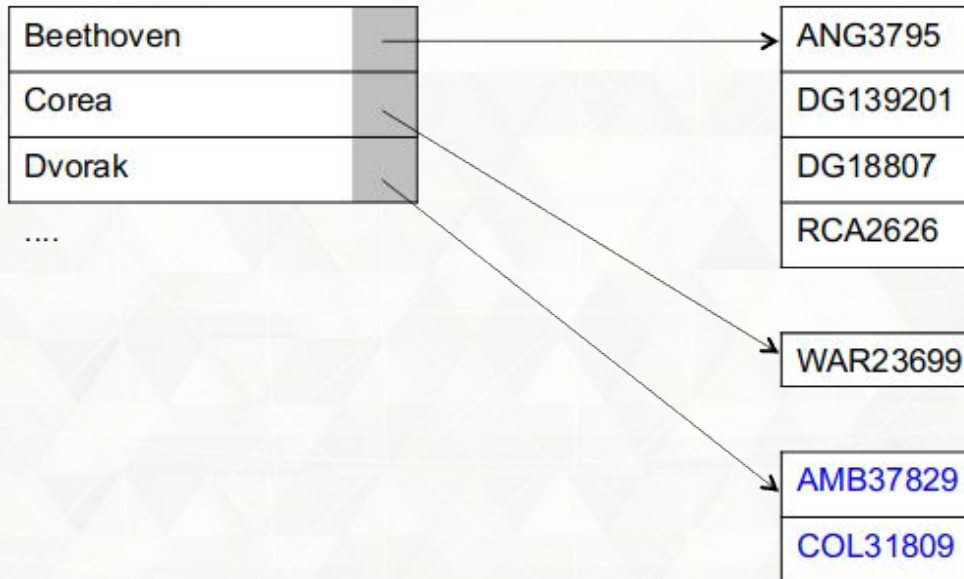
- Associa uma lista encadeada das chaves primárias a cada chave secundária
- Visão Conceitual



- E se inserimos um novo registro relativo a Dvorak?

Listas Invertidas

- Inserção de um novo registro relativo a Dvorak
- Visão Conceitual



Listas Invertidas - Implementação

índice secundário

Beethoven	3
Corea	2
Dvorak	7
Prokofiev	0
Rimsky	6
Springsteen	4
Sweet Honey In The	9

....

lista invertida

0	LON2312	-1
1	RCA2626	-1
2	WAR23699	-1
3	ANG3795	8
4	COL38358	-1
5	DG18807	1
6	MER75016	-1
7	COL31809	-1
8	DG139201	5
9	FF245	-1

aquivo de dados

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

*RRN da primeira referencia
da chave secundária*

*RRN da próxima referencia da chave secundária na lista
invertida ou -1*

Listas Invertidas - Implementação

- Inserção de um novo registro relativo a Dvorak

índice secundário

Beethoven	3
Corea	2
Dvorak	10
Prokofiev	0
Rimsky	6
Springsteen	4
Sweet Honey In The	9

lista invertida

0	LON2312	-1
1	RCA2626	-1
2	WAR23699	-1
3	ANG3795	8
4	COL38358	-1
5	DG18807	1
6	MER75016	-1
7	COL31809	-1
8	DG139201	5
9	FF245	-1
10	AMB37829	7

aquivo de dados

32	LON 2312 Rom ...
77	RCA 2626 Quar ...
132	WAR 23699 Tou ...
167	ANG 3795 Sym ...
211	COL 38358 Neb ...
256	DG 18807 Sym ...
300	MER 75016 Coq ...
353	COL 31809 Sym ...
396	DG 139201 Violi ...
442	FF 245 Good ...
530	AMB 37829 Bye ...

Listas Invertidas - Vantagens

- Índice secundário
 - alterado quando insere-se um registro com chave inexistente, ou quando altera-se chave já existente
- Remoção, inserção ou alteração de registros já existentes
 - alteração apenas no arquivo da lista invertida
 - modificação do campo de referência do índice se necessário
- Ordenação do arquivo de índice secundário
 - mais rápida: menos registros e registros menores
- Registros de tamanho fixo
 - facilita a adoção de um mecanismo para reaproveitamento de espaço

Listas Invertidas - Desvantagens

- Chaves primárias associadas a uma certa chave secundária não estão adjacentes fisicamente no disco
 - pode ser necessário realizar vários seeks para recuperar os valores na lista invertida
- Ideal
 - Manter o índice e a lista na memória primária