

Engenharia de Computação

Estrutura de Dados 2

Aula 11 – Indexação Primária

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

Índice

- Em geral, um índice **fornece mecanismos para localizar informações**
 - Índice de um livro ou catálogo de uma biblioteca
 - Facilita muito o trabalho de busca!
- Em arquivos
 - Permite localizar registros rapidamente
 - Não é necessário ordenar arquivo de dados, nem quando novos registros são adicionados

Índice

- Índices Simples
 - Representados em arrays , cuja estrutura contém as chaves e os campos de referência
- Exemplo: uma enorme coleção de CDs
 - Registros de tamanho variável
 - ID Number: Número de identificação
 - Title: Título
 - Composer: Compositor(es)
 - Artist: Artista(s)
 - Label: Rótulo (código da gravadora)

Índice Simples

- Chave primária: combinação de Label + ID Number
 - Poderia ser qualquer outro campo ou combinação de campos que fosse único para cada registro

Record address	ID		Title	Composer(s)	Artist(s)
	Label	number			
17	LON	2312	Romeo and Juliet	Prokofiev	Maazel
62	RCA	2626	Quartet in C Sharp Minor	Beethoven	Julliard
117	WAR	23699	Touchstone	Corea	Corea
152	ANG	3795	Symphony No. 9	Beethoven	Giulini
196	COL	38358	Nebraska	Springsteen	Springsteen
241	DG	18807	Symphony No. 9	Beethoven	Karajan
285	MER	75016	Coq d'Or Suite	Rimsky-Korsakov	Leinsdorf
338	COL	31809	Symphony No. 9	Dvorak	Bernstein
382	DG	139201	Violin Concerto	Beethoven	Ferras
427	FF	245	Good News	Sweet Honey in the Rock	Sweet Honey in the Rock

Índice Simples

Index		Recording file	
Key	Reference field	Address of record	Actual data record
ANG3795	152	17	LON 2312 Romeo and Juliet Prokofiev ...
COL31809	338	62	RCA 2626 Quartet in C Sharp Minor Beethoven ...
COL38358	196	117	WAR 23699 Touchstone Corea ...
DG139201	382	152	ANG 3795 Symphony No. 9 Beethoven ...
DG18807	241	196	COL 38358 Nebraska Springsteen ...
FF245	427	241	DG 18807 Symphony No. 9 Beethoven ...
LON2312	17	285	MER 75016 Coq d'Or Suite Rimsky-Korsakov ...
MER75016	285	338	COL 31809 Symphony No. 9 Dvorak ...
RCA2626	62	382	DG 139201 Violin Concerto Beethoven ...
WAR23699	117	427	FF 245 Good News Sweet Honey in the Rock ...

Índice Simples

- O índice consiste, em geral, em um outro arquivo com registros de tamanho fixo
 - Mesmo que o arquivo principal com os dados não tenha registros de tamanho fixo
- Cada registro do índice contém pelo menos 2 campos de tamanho fixo
 - Chave
 - Posição inicial (byte offset) ou RRN do registro no arquivo de dados
- A cada registro do arquivo de dados corresponde a um registro no índice
 - O índice está ordenado, apesar do arquivo de dados não estar
 - Em geral, o arquivo de dados está organizado segundo a ordem de entrada dos registros (*entry sequenced file*)

Índice Simples

- Vantagens do arquivo de índice sobre o de dados
 - Mais fácil de trabalhar, pois usa registros de tamanho fixo
 - Pode ser pesquisado com busca binária (em memória principal, inclusive, se valer a pena carregá-lo)
 - É muito menor do que o arquivo de dados
- Registros de tamanho fixo no arquivo índice impõem um limite ao tamanho da chave primária
- Os registros do índice poderiam conter outros campos além da chave/ offset (por exemplo, o tamanho do registro)

Índice Simples

- Como são feitas as operações básicas num arquivo **indexado** e “**entry-sequenced**”?
 - Inserção de registros
 - No arquivo de dados: no final;
 - Inserção de novo registro de chave no índice
 - Remoção de registros
 - Busca do registro usando o índice: 1 seek + remoção
 - Remoção de registro de chave no índice
 - Atualização de registros
 - Busca no índice + seek
 - Se mudar chave, alterar também o índice
 - Busca de registros
 - Busca no índice + seek

Índice Simples

- Para índices que cabem em memória
 - Criar arquivos índice e de dados
 - Carregar índice para memória
 - Inserir registro
 - Inserção deve ser feita no arquivo de dados...
 - e também no índice, que eventualmente será reorganizado
 - Eliminar registro
 - Remove do arquivo de dados (aula anterior)
 - Remove também do índice
 - Índice pode ser reorganizado ou pode-se apenas marcar os registros excluídos

Índice Simples

- Atualizar registro - duas categorias
 - Muda o valor da chave
 - Muda o conteúdo do registro
- Atualizar índice no disco: caso sua cópia em memória tenha sido alterada
 - É obrigatório que o programa se proteja contra índices desatualizados
- Possíveis estratégias?

Como evitar índices desatualizados

- Deve haver um mecanismo que permita saber se o índice está atualizado em relação ao arquivo de dados
 - Possibilidade: um *status flag* é setado no arquivo índice mantido em disco assim que a sua cópia na memória é alterada
 - Esse flag pode ser mantido no *registro header* do arquivo índice, e atualizado sempre que o índice é reescrito no disco
 - Se um programa detecta que o índice está desatualizado, uma função deve ser ativada para **reconstruir o índice** a partir do arquivo de dados

Índices muito grandes

- Se o índice **não cabe inteiro na memória**, o acesso e manutenção precisam ser feitos em memória secundária
- Não é mais aconselhável usar índices simples, uma vez que:
 - A busca binária pode exigir vários acessos a disco
 - A necessidade de deslocar registros nas inserções e remoções de registros tornaria a manutenção do índice excessivamente cara

Índices muito grandes

- Utilizam-se outras organizações
 - **Hashing**, caso a velocidade de acesso seja a prioridade máxima
 - Acesso direto apenas
 - **Árvores-B**, caso se deseje combinar acesso por chaves e acesso sequencial eficientemente

Índices muito grandes

- Ainda assim, o uso de índice simples que não cabe em RAM ainda é mais vantajoso do que manter um arquivo de dados ordenado e não indexado.
 - Tem registros de tamanho fixo, portanto, é possível fazer busca binária;
 - É bem menor do que arquivo de dados, facilitando a manutenção;
 - Índices fornecem múltiplas visões de um conjunto de dados