

Classificação e Pesquisa de Dados

Aula 23

Organização de Arquivos: Arquivos Indexados, Diretos e Arquivos Invertidos

UFRGS

INF01124

Arquivo indexado - AI

- ◆ Indicação de uso
 - Baixa frequência de acessos sequenciais em relação aos acessos aleatórios

Arquivo indexado

- ◆ Caracterização
 - Nesta organização, os registros são acessados, sempre, através de um ou mais índices, não havendo qualquer compromisso com a ordem física de instalação dos registros no arquivo
 - Consequência direta → maior eficiência na operação de inserção de um registro

Arquivo Indexado

chave		endereço	
1000	5	#1	1075
1050	7	#2	1405
1075	1	#3	1100
		#4	1510
1100	3	#5	1000
1300	12	#6	1600
1350	10	#7	1050
		#8	1480
1400	9	#9	1400
1405	2	#10	1350
1480	8	#11	1800
...		#12	1300
	

Tabela Fina

Índices

- ◆ Podem existir tantas quantas forem as chaves de acesso aos registros
- ◆ Um índice contém uma entrada para cada registro

Índices

- ◆ As entradas do índice são ordenadas pelo valor da chave de acesso
- ◆ Cada entrada em um índice é constituída por um par
 - ♦ <Chave_registro, endereço_registro>

Tipos de Índices

- ◆ **Exaustivo**
 - Contém uma entrada para cada registro do arquivo
- ◆ **Seletivo**
 - Possui entradas apenas para um subconjunto dos registros
 - Ex.: Subconjunto que define todos os alunos cujo curso = "ciência da computação" e ano_ingresso = "2006"

Desvantagens

- ◆ Necessidade de atualização de todos os índices quando
 - Um registro é inserido no arquivo
 - Alteração de um registro envolve atributos associados a índices
- ◆ Quanto maior é o número de índices, mais onerosa é a operação de inserção

Inserção de um registro

- ◆ O registro é armazenado em qualquer endereço vago do arquivo
- ◆ Seus pares <chave_registro, endereço_registro> relativos aos índices existentes são inseridos nos índices correspondentes

Exclusão de um registro

- ◆ Área de dados ocupada é liberada
- ◆ São removidas as entradas do índices correspondentes por
 - Retirada física do par
 - Utilização de uma marca
- ◆ Devem ser determinados os índices seletivos afetados

Alteração de um registro

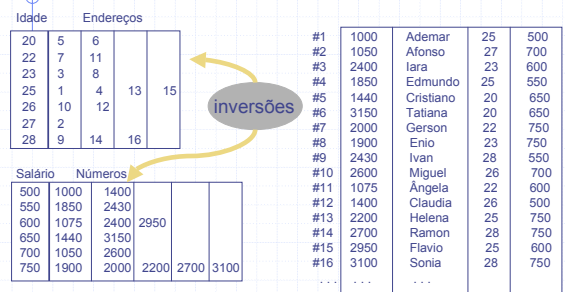
- ◆ **Identificação do registro**
 - Via argumento de pesquisa
 - Endereço é determinado por uma busca sobre o índice
 - O endereço do registro é conhecido
- ◆ **Tipo da alteração**
 - Normal
 - Lê → altera → grava na mesma posição
 - Com aumento do tamanho do registro
 - Lê → exclui → insere registro atualizado

Arquivo invertido

◆ Caracterização

- Em vez de serem coletados os valores dos atributos para cada registro, são identificados os registros que possuem um dado valor do atributo considerado
- À cada valor de chave corresponde uma lista de endereços de registros
- O conjunto de listas invertidas associado a uma chave de acesso é chamado ***inversão***
- Um arquivo invertido pode possuir uma ou mais inversões

Arquivo Invertido



Arquivo invertido

◆ Indicação de uso

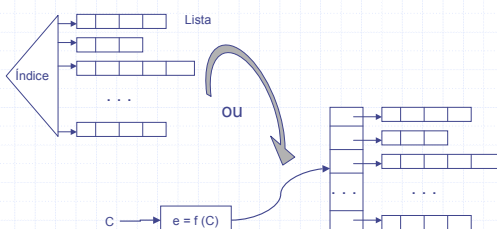
- Esta organização está voltada para o acesso aos registros de um arquivo por chaves secundárias

Arquivo Invertido

◆ Decisões importantes

- Como estruturar o acesso às listas?
- Como estruturar as listas?

Como estruturar o acesso às listas?



Como estruturar as listas

- ◆ Qualquer solução estudada para representação de listas lineares
 - Contigüidade física --> registros de tamanho variável, normalmente
 - Encadeamento
 - Mapa de bits (quando a gama de valores possíveis é pequena)
- ◆ Considerar que, via de regra, tais listas são armazenadas em disco, não sendo recomendável o simples encadeamento item a item

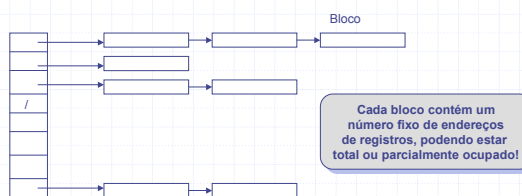
Como estruturar as listas

- ◆ Solução usual: fazer com que cada lista seja composta por uma lista encadeada de zero ou mais blocos, cada um contendo vários endereços de registros
- ◆ Fazer com que as listas estejam ordenadas com o mesmo critério

Mapa de bits

	500	550	600	650	700	750					
#1	1	0	0	0	0	0	#1	1000	Ademar	25	500
#2	0	0	0	0	1	0	#2	1050	Alfonso	27	700
#3	0	0	0	1	0	0	#3	2400	Iara	23	600
#4	0	1	0	0	0	0	#4	1850	Edmundo	25	550
#5	0	0	0	1	0	0	#5	1440	Cristiano	20	650
#6	0	0	0	0	1	0	#6	3150	Tatiana	20	650
#7	0	0	0	0	0	0	#7	2000	Gerson	22	750
#8	0	0	0	0	0	1	#8	1900	Enio	23	750
#9	0	1	0	0	0	0	#9	2430	Ivan	28	550
#10	0	0	0	0	1	0	#10	2600	Miguel	26	700
#11	0	0	1	0	0	0	#11	1075	Ângela	22	600
#12	1	0	0	0	0	0	#12	1400	Claudia	26	500
#13	0	0	0	0	0	1	#13	2200	Helena	25	750
#14	0	0	0	0	0	1	#14	2700	Ramon	28	750
#15	0	0	1	0	0	0	#15	2950	Flavio	25	600
#16	0	0	0	0	0	1	#16	3100	Sonia	28	750
									

Como estruturar as listas



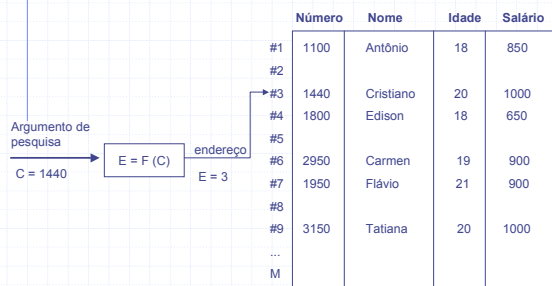
Arquivo direto (*hash*)

- ◆ Indicação de uso
 - Baixa frequência de acessos seriais em relação aos acessos aleatórios
 - Dispositivos de armazenamento de acesso direto

Arquivo direto

- ◆ Caracterização
 - Organização por cálculo de endereço, com área de dados organizadas em blocos
 - Instalação dos registros em endereços determinados, com base no valor de uma chave primária
 - Acesso aleatório eficiente
 - Dispensa uso de estrutura auxiliar (índice)
 - Substituição de um índice pela função que calcula o endereço do registro a partir do argumento de busca

Arquivo direto



Tipos de funções

- ◆ Funções determinísticas
 - Um único valor de chave de acesso para cada endereço
 - Impossível na prática
- ◆ Funções probabilísticas
 - Geram para cada valor da chave um endereço "tão único quanto possível"
 - Geram a necessidade de tratamento das colisões

Tratamento das colisões

- ◆ Endereçamento aberto
 - Pesquisa seqüencial
 - Pesquisa no bloco
 - Realeatorização
- ◆ Encadeamento
 - Uso de áreas de extensão
 - Encadeamento puro