

Ordenação de Arquivos



Prof. Flávio Humberto Cabral Nunes

Conteúdo

1. Introdução
 2. Intercalação Balanceada de Vários Caminhos
 3. Implementação através de Seleção por Substituição
 4. Considerações Práticas
- Capítulo: 5 (APOSTILA).

Introdução

- Ordenação externa é usada quando o número de registros é maior do que a memória RAM
- Os métodos de ordenação externa são diferentes dos métodos de ordenação interna
 - Mas o problema é o mesmo
- Deve-se considerar que os dados estão em unidades de armazenamento secundário

Introdução

- Nas memórias externas, os dados são armazenados em um arquivo sequencial
 - Apenas um registro é acessado por vez
- Diferente da estrutura de dados vetor
- Portanto são necessárias novas técnicas para ordenação de registros

Por que ordenação externa é diferente da ordenação interna?

1. Custo de acesso a memória secundária é muito maior do que o acesso a memória primária
2. Restrições de acesso a dados
 - Fitas são acessadas somente sequencialmente
 - Em discos, acesso direto é muito caro
3. Os métodos de ordenação externa são dependentes do estado atual da tecnologia

Ordenação Externa

- Dá-se ênfase na minimização do número de vezes que cada item é transferido entre memória interna e externa
- Transferência deve ser tão eficiente quanto os equipamentos permitam
- Método mais importante é a intercalação
 - Combinação de 2 ou mais blocos ordenados para produzir um único bloco ordenado

Estratégia Geral dos Métodos de Ordenação Externa

1. Arquivo é quebrado em blocos do tamanho da memória interna disponível e cada bloco é ordenado
2. Os blocos são ordenados fazendo-se várias passadas no arquivo. A cada passada, blocos maiores são criados até que todo arquivo esteja ordenado

Eficiência

- Os algoritmos devem reduzir o número de passadas sobre o arquivo
- Maior custo está nas operações de entrada e saída
- Medida de complexidade = no. de vezes que um item é lido e escrito na memória auxiliar
- Bons métodos fazem menos de 10 passadas

Intercalação Balanceada de Vários Caminhos

- Considere que o arquivo está armazenado em fita magnética
- Memória interna só tem espaço para 3 registros
- Têm-se 6 unidades de fita magnética
- Conteúdo do arquivo:
 - INTERCALACAOBALANCEADA

Intercalação Balanceada de Vários Caminhos

Memória: ERC

FITA 1: INT ACO ADE

FITA 2: CER ABL A

FITA 3: AAL ACN

FITA 4:

FITA 5:

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: IAC
↑↑↑

FITA 1:	I	N T	A C O	A D E
FITA 2:	C	E R	A B L	A
FITA 3:	A	A L	A C N	

FITA 4: A

FITA 5:

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: IAC



FITA 1: ~~I~~ N T A C O A D E

FITA 2: ~~C~~ E R A B L A

FITA 3: ~~A~~ ~~A~~ L A C N

FITA 4: A A

FITA 5:

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: ILC



FITA 1: ~~I~~ N T A C O A D E

FITA 2: ~~C~~ E R A B L A

FITA 3: ~~A~~ ~~A~~ ~~L~~ A C N

FITA INATIVA

FITA 4: A A C

FITA 5:

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: ILE



FITA 1: ~~I~~ N T A C O A D E

FITA 2: ~~C~~ ~~E~~ R A B L A

FITA 3: ~~A~~ ~~A~~ ~~L~~ A C N

FITA INATIVA

FITA 4: A A C E

FITA 5:

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: ILR



FITA 1: ~~I~~N T A C O A D E

FITA 2: ~~C~~~~E~~~~R~~ A B L A FITA INATIVA

FITA 3: ~~A~~~~A~~~~L~~ A C N FITA INATIVA

FITA 4: A A C E I

FITA 5:

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: NLR



FITA 1: ~~IN~~T A C O A D E

FITA 2: ~~CEP~~ A B L A FITA INATIVA

FITA 3: ~~AAA~~ A C N FITA INATIVA

FITA 4: A A C E I L

FITA 5:

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: **N**T R



FITA 1: ~~INT~~ A C O A D E

FITA 2: ~~CEP~~ A B L A

FITA 3: ~~AAE~~ A C N

FITA 4: A A C E I L **N R T**

FITA 5:

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: AAA



FITA 1: ~~INT~~ A C O A D E

FITA 2: ~~CEP~~ A B L A

FITA 3: ~~AAL~~ A C N

FITA 4: A A C E I L N R T

FITA 5: A

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: CAA



FITA 1: ~~INT~~ ~~ACO~~ A D E

FITA 2: ~~CEP~~ ~~ABL~~ A

FITA 3: ~~AA~~ ~~ACN~~

FITA 4: A A C E I L N R T

FITA 5: A A

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: CBA



FITA 1: ~~INT~~ ~~ACO~~ A D E

FITA 2: ~~CEP~~ ~~ABL~~ A

FITA 3: ~~AA~~ ~~AC~~ N

FITA 4: A A C E I L N R T

FITA 5: A A A

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: CBC



FITA 1: ~~INT~~ ~~ACO~~ A D E

FITA 2: ~~CEP~~ ~~ABL~~ A

FITA 3: ~~AAL~~ ~~ACN~~

FITA 4: A A C E I L N R T

FITA 5: A A A B

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: CLC



FITA 1: ~~INT~~ ~~INT~~ ~~AC~~ ~~O~~ A D E

FITA 2: ~~CEP~~ ~~AB~~ ~~BL~~ A

FITA INATIVA

FITA 3: ~~AA~~ ~~AL~~ ~~AC~~ N

FITA 4: A A C E I L N R T

FITA 5: A A A B C

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: OLC



FITA 1: ~~INT~~ ~~ACC~~ A D E FITA INATIVA

FITA 2: ~~CEP~~ ~~ABL~~ A FITA INATIVA

FITA 3: ~~AAL~~ ~~AC~~ N

FITA 4: A A C E I L N R T

FITA 5: A A A B C C

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: OLN



FITA 1: ~~INT~~ ~~ACC~~ A D E

FITA 2: ~~CEP~~ ~~ABL~~ A

FITA 3: ~~AA~~ ~~ACN~~

FITA 4: A A C E I L N R T

FITA 5: A A A B C C L N O

FITA 6:

Intercalação Balanceada de Vários Caminhos

Memória: AAD



FITA 1: ~~INT~~ ~~ACC~~ ~~AD~~ E

FITA 2: ~~CEP~~ ~~ABL~~ ~~A~~

FITA 3: ~~AA~~ ~~ACN~~

FITA INATIVA

FITA 4: A A C E I L N R T

FITA 5: A A A B C C L N O

FITA 6: A

Intercalação Balanceada de Vários Caminhos

Memória: EAD



FITA 1: ~~INT~~ ~~ACC~~ ~~ADE~~

FITA INATIVA

FITA 2: ~~CEP~~ ~~ABL~~ ~~A~~

FITA INATIVA

FITA 3: ~~AA~~ ~~ACN~~

FITA 4: A A C E I L N R T

FITA 5: A A A B C C L N O

FITA 6: A A D E

Intercalação Balanceada de Vários Caminhos

Memória: AAA
↑↑↑

FITA 1: A

FITA 2:

FITA 3:

FITA 4: ~~A~~ A C E I L N R T

FITA 5: ~~A~~ A A B C C L N O

FITA 6: ~~A~~ A D E

Intercalação Balanceada de Vários Caminhos

Memória: AAA



FITA 1: A A

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~ C E I L N R T

FITA 5: ~~A~~ A A B C C L N O

FITA 6: ~~A~~ A D E

Intercalação Balanceada de Vários Caminhos

Memória: CAA



FITA 1: A A A

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~ E I L N R T

FITA 5: ~~A~~ A A B C C L N O

FITA 6: ~~A~~ A D E

Intercalação Balanceada de Vários Caminhos

Memória: CAA



FITA 1: A A A A

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~ E I L N R T

FITA 5: ~~A~~~~A~~ A B C C L N O

FITA 6: ~~A~~ A D E

Intercalação Balanceada de Vários Caminhos

Memória: CAA



FITA 1: A A A A A

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~ E I L N R T

FITA 5: ~~A~~~~A~~~~A~~ B C C L N O

FITA 6: ~~A~~ A D E

Intercalação Balanceada de Vários Caminhos

Memória: CBA



FITA 1: A A A A A A

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~ E I L N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~ C C L N O

FITA 6: ~~A~~ A D E

Intercalação Balanceada de Vários Caminhos

Memória: CBA



FITA 1: A A A A A A A

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~ E I L N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~ C C L N O

FITA 6: ~~A~~~~A~~ D E

Intercalação Balanceada de Vários Caminhos

Memória: CBD



FITA 1: A A A A A A A B

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~ E I L N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~ C C L N O

FITA 6: ~~A~~~~A~~~~D~~ E

Intercalação Balanceada de Vários Caminhos

Memória: CCD



FITA 1: A A A A A A A B C

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~ E I L N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~~~C~~ C L N O

FITA 6: ~~A~~~~A~~~~D~~ E

Intercalação Balanceada de Vários Caminhos

Memória: ECD



FITA 1: A A A A A A A B C C

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~~~E~~ I L N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~~~C~~ C L N O

FITA 6: ~~A~~~~A~~~~D~~ E

Intercalação Balanceada de Vários Caminhos

Memória: ECD



FITA 1: A A A A A A A B C C C

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~~~E~~ I L N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~~~C~~~~C~~ L N O

FITA 6: ~~A~~~~A~~~~D~~ E

Intercalação Balanceada de Vários Caminhos

Memória: ELD



FITA 1: A A A A A A A B C C C D

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~~~E~~ I L N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~~~C~~~~C~~~~E~~ N O

FITA 6: ~~A~~~~A~~~~D~~ E

Intercalação Balanceada de Vários Caminhos

Memória: ELE



FITA 1: A A A A A A A B C C C D E

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~~~E~~ I L N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~~~C~~~~C~~~~C~~ N O

FITA 6: ~~A~~~~A~~~~D~~~~E~~

FITA INATIVA

Intercalação Balanceada de Vários Caminhos

Memória: ILE



FITA 1: A A A A A A A B C C C D E E

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~~~E~~~~L~~ N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~~~C~~~~C~~~~C~~ N O

FITA 6: ~~A~~~~A~~~~D~~~~E~~

FITA INATIVA

Intercalação Balanceada de Vários Caminhos

Memória: ILL



FITA 1: A A A A A A A B C C C D E E I

FITA 2:

FITA 3:

FITA 4: ~~A~~~~A~~~~C~~~~E~~~~L~~~~L~~ N R T

FITA 5: ~~A~~~~A~~~~A~~~~B~~~~C~~~~C~~~~C~~ N O

FITA 6: ~~A~~~~A~~~~D~~~~E~~

FITA INATIVA

Intercalação Balanceada de Vários Caminhos

Memória: NLL



FITA 1: A A A A A A A B C C C D E E I L

FITA 2:

FITA 3:

FITA 4: ~~A~~ ~~A~~ ~~C~~ ~~E~~ ~~L~~ ~~L~~ ~~N~~ R T

FITA 5: ~~A~~ ~~A~~ ~~A~~ ~~B~~ ~~C~~ ~~C~~ ~~C~~ N O

FITA 6: ~~A~~ ~~A~~ ~~D~~ ~~E~~

FITA INATIVA

Intercalação Balanceada de Vários Caminhos

Memória: NRL



FITA 1: A A A A A A A B C C C D E E I L L

FITA 2:

FITA 3:

FITA 4: ~~A~~ ~~A~~ ~~C~~ ~~E~~ ~~L~~ ~~L~~ ~~N~~ ~~R~~ T

FITA 5: ~~A~~ ~~A~~ ~~A~~ ~~B~~ ~~C~~ ~~C~~ ~~C~~ ~~N~~ O

FITA 6: ~~A~~ ~~A~~ ~~D~~ ~~E~~

FITA INATIVA

Intercalação Balanceada de Vários Caminhos

Memória: NRN



FITA 1: A A A A A A A B C C C D E E I L L N

FITA 2:

FITA 3:

FITA 4: ~~A~~ ~~A~~ ~~C~~ ~~E~~ ~~L~~ ~~L~~ ~~N~~ ~~R~~ T

FITA 5: ~~A~~ ~~A~~ ~~B~~ ~~C~~ ~~C~~ ~~C~~ ~~N~~ ~~N~~ O

FITA 6: ~~A~~ ~~A~~ ~~D~~ ~~E~~

FITA INATIVA

FITA INATIVA

Intercalação Balanceada de Vários Caminhos

Memória: NRO



FITA 1: A A A A A A A B C C C D E E I L L N N

FITA 2:

FITA 3:

FITA 4: ~~A~~ ~~A~~ ~~C~~ ~~E~~ ~~L~~ ~~L~~ ~~N~~ ~~N~~ T

FITA 5: ~~A~~ ~~A~~ ~~B~~ ~~C~~ ~~C~~ ~~C~~ ~~N~~ ~~N~~ ~~O~~

FITA 6: ~~A~~ ~~A~~ ~~D~~ ~~E~~

FITA INATIVA

FITA INATIVA

Intercalação Balanceada de Vários Caminhos

Memória: TRO



FITA 1: A A A A A A A B C C C D E E I L L N N O R T

FITA 2:

FITA 3:

FITA 4: ~~A~~ ~~A~~ ~~C~~ ~~E~~ ~~L~~ ~~L~~ ~~N~~ ~~N~~ ~~O~~ ~~R~~ ~~T~~

FITA 5: ~~A~~ ~~A~~ ~~B~~ ~~C~~ ~~C~~ ~~C~~ ~~D~~ ~~E~~ ~~E~~ ~~I~~ ~~L~~ ~~L~~ ~~N~~ ~~N~~ ~~O~~ ~~R~~ ~~T~~

FITA 6: ~~A~~ ~~A~~ ~~D~~ ~~E~~

FITA INATIVA

FITA INATIVA

Quantas passadas são necessárias?

- Considere um arquivo com n registros (um registro é igual a uma palavra) e uma memória interna de m palavras
- Passada inicial produz n/m blocos ordenados
- Se cada registro tiver k palavras, $k > 1$
- Passada inicial produz $n/m/k$ blocos

Quantas passadas são necessárias?

- Seja a função de complexidade P tal que
 - $P(x)$ seja o número de passadas
 - f o número de fitas
- Logo
 - $P(n) = \log_f n/m$
- No exemplo, $n = 22$, $m = 3$ e $f = 6$
 - $P(22) = \log_6 22/3 = 2$

Quantas passadas são necessárias?

- No exemplo foram utilizadas $2f$ fitas
- Para usar $f + 1$ fitas
 - Encaminhe todos os blocos para uma fita
 - Redistribua estes blocos entre as fitas de onde eles foram lidos
- Possui o custo de mais passadas

Implementação através de Seleção por Substituição

- A intercalação pode ser feita utilizando **filas de prioridades**
- Tanto a quebra do arquivo em blocos como a intercalação são implementadas de forma eficiente
- Substitui o menor registro existente na memória interna pelo próximo registro da fita de entrada
- Implementada através de um *heap*

Implementação através de Seleção por Substituição

- Processo começa fazendo m inserções na fila de prioridades vazia
- O menor item da fila de prioridades é substituído pelo próximo item de entrada
 - Se o próximo item é menor do que o que está saindo, então ele deve ser marcado como membro do próximo bloco e tratado como maior do que todos os itens do bloco corrente
- Quando um item marcado vai para o topo da fila, o bloco corrente é encerrado e novo bloco ordenado é iniciado

Implementação através de Seleção por Substituição

- Considere um heap de tamanho 3
- A condição do heap é que a primeira chave tem que ser menor do que a segunda e a terceira
- Blocos gerados
 - INRT – tamanho 4
 - ACEL – tamanho 4
 - AABCLO – tamanho 6
 - AACEN – tamanho 5
 - AAD – tamanho 6

Implementação através de Seleção por Substituição

- Após encontrados os blocos, faz-se a intercalação dos mesmos
- A intercalação também pode ser feita utilizando fila de prioridades

Considerações Práticas

- Operações de entrada e saída de dados devem ser implementadas de forma eficiente
- Deve-se procurar fazer leitura, escrita e processamento de forma simultânea
- Os computadores de maior porte possuem uma ou mais unidades independentes para processamento de E/S

Considerações Práticas

- Técnica para obter superposição de E/S e processamento interno:
 - Utilize 2f áreas de entrada e 2f de saída
 - Para cada unidade de entrada ou saída, utiliza-se duas áreas de armazenamento
 - Uma para uso do processador central
 - Outra para uso do processador de entrada ou saída

Considerações Práticas

- Para entrada, o processador central usa uma das duas áreas enquanto a unidade de entrada está preenchendo a outra área
- Depois a utilização das áreas é invertida entre o processador de entrada e o processador central
- Para saída, a mesma técnica é utilizada

Considerações Práticas

- Problemas:
 - Apenas metade da memória disponível é utilizada
 - Isso pode levar a uma ineficiência se o número de áreas for grande
 - Todas as f áreas de entrada em uma intercalação de f caminhos se esvaziando aproximadamente ao mesmo tempo

Considerações Práticas

- Solução: Técnica de previsão
- Requer a utilização de uma única área extra de armazenamento durante a intercalação
- Superpõe a entrada da próxima área que precisa ser preenchida com a parte de processamento interno do algoritmo
- A área cujo último registro é o menor será a primeira a se esvaziar