

INF 112: Programação II

Aula 10

→ Algoritmos de ordenação – parte 4 Ordenação externa

Aula baseada no livro: *Projetos de Algoritmos com Implementações em Pascal e C*, Nivio Ziviani, 4ed.

Algoritmos de ordenação – Ordenação Externa



- A ordenação externa envolve entrada composta por um número de registros que é maior do que a memória interna do computador pode armazenar.
- Na ordenação externa as estruturas de dados têm que levar em conta o fato de que os dados estão armazenados em unidades de memória externa, relativamente muito mais lentas do que a memória principal.
- Nas memórias externas como as fitas magnéticas, os dados são armazenados como um arquivo sequencial, onde apenas um registro pode ser acessado em um dado momento. Esta é uma restrição forte se comparada com as possibilidades de acesso da estrutura de dados do tipo arranjo.



Algoritmos de ordenação – Ordenação Externa



- Os itens armazenados em disco magnético podem ser acessados diretamente, mas a um custo maior do que o custo para acessar sequencialmente, o que contra-indica o uso do acesso direto.
- Consequentemente, os métodos de ordenação interna apresentados até agora são inadequados para ordenação externa.
- Além de direcionar a ordenação para fazer leituras sequenciais, há que se dar ênfase na minimização do número de vezes que cada item é transferido entre a memória interna e a memória externa.



Algoritmos de ordenação – Ordenação Externa



- ➔ O método de ordenação externa mais importante é o método de ordenação por intercalação (merge).
- ➔ Intercalar significa combinar dois ou mais blocos ordenados em um único bloco ordenado através de seleções repetidas entre os itens disponíveis em cada momento.
- ➔ A maioria dos métodos de ordenação externa utilizam a seguinte estratégia geral:
 1. É realizada uma primeira passada sobre o arquivo, quebrando-o em blocos do tamanho da memória interna disponível. Cada bloco é então ordenado na memória interna.
 2. Os blocos ordenados são intercalados, fazendo várias passadas sobre o arquivo. A cada passada são criados blocos ordenados cada vez maiores, até que todo o arquivo esteja ordenado.



Algoritmos de ordenação – Ordenação Externa



- Vamos considerar o processo de ordenação externa quando o arquivo está armazenado em fita magnética.
- Para apresentar os vários passos envolvidos em um algoritmo de ordenação por intercalação balanceada vamos utilizar um arquivo exemplo.
- Considere um arquivo armazenado em uma fita de entrada composto pelos registros com as chaves mostradas na figura abaixo. Os 22 registros devem ser ordenados de acordo com as chaves e colocados em uma fita de saída. Neste caso os registros são lidos sequencialmente.

I N T E R C A L A C A O B A L A N C E A D A

Arquivo exemplo com 22 registros



Algoritmos de ordenação – Ordenação Externa

- Assuma que a memória interna do computador a ser utilizado só tem espaço para três registros e o número de unidades de fita magnética é seis.
- Na primeira etapa o arquivo é lido de três em três registros. Cada bloco de três registros é ordenado e escrito em uma das fitas de saída.
- Para o exemplo mostrado na figura anterior, são lidos os registros INT e escrito o bloco INT na fita 1. A seguir são lidos os registros ERC e escrito o bloco CER na fita 2 e assim por diante, conforme ilustra a figura abaixo. Três fitas são utilizadas em uma intercalação-de-3-caminhos.

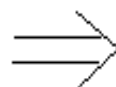
fita 1:	<i>I N T</i>	<i>A C O</i>	<i>A D E</i>
fita 2:	<i>C E R</i>	<i>A B L</i>	<i>A</i>
fita 3:	<i>A A L</i>	<i>A C N</i>	

Formação dos blocos ordenados iniciais

Algoritmos de ordenação – Ordenação Externa

- Na segunda etapa os blocos ordenados devem ser intercalados. O primeiro registro de cada uma das três fitas é lido para a memória interna, ocupando-a por inteiro. A seguir, o registro contendo a menor chave dentre as três é retirado (e colocado em uma fita de saída) e o próximo registro da mesma fita é lido para a memória interna, repetindo-se o processo.
- Quando o terceiro registro de um dos blocos é lido aquela fita fica inativa até que o terceiro registro das outras fitas também sejam lidos e escritos na fita de saída, formando um bloco de 9 registros ordenados. A seguir, o segundo bloco de 3 registros de cada fita é lido para formar outro bloco ordenado de nove registros, o qual é escrito em uma outra fita. Ao final, três novos blocos ordenados são obtidos, conforme mostra a figura.

fitas 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*



fitas 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-de-3-caminhos

Algoritmos de ordenação – Ordenação Externa



- Em seguida, mais uma intercalação-de-3-caminhos das fitas 4, 5 e 6 para as fitas 1, 2 e 3 completa a ordenação.
- Se o arquivo exemplo tivesse um número maior de registros, então vários blocos ordenados de 9 registros seriam formados nas fitas 4, 5 e 6. Neste caso, a segunda passada produziria blocos ordenados de 27 registros nas fitas 1, 2 e 3; a terceira passada produziria blocos ordenados de 81 registros nas fitas 4, 5 e 6, e assim sucessivamente, até obter-se um único bloco ordenado.
- Neste ponto cabe a seguinte pergunta:
 - Quantas passadas são necessárias para ordenar um arquivo de tamanho arbitrário?



Algoritmos de ordenação – Ordenação Externa



- Considere um arquivo contendo n registros (neste caso cada registro contém apenas uma *palavra*) e uma memória interna de m palavras. A passada inicial sobre o arquivo produz n/m blocos ordenados.
- Podemos considerar como operação fundamental o número de passadas, que chamaremos de $P(n)$, para realização da intercalação dos blocos ordenados.
- Seja ainda f o número de fitas utilizadas em cada passada. Para uma intercalação-de- f -caminhos o número de passadas é:

$$P(n) = \log_f n/m$$

No exemplo mostrado $n=22$, $m=3$ e $f=3$, logo:

$$P(n) = \log_3 22/3 \Rightarrow 2.$$



Algoritmos de ordenação – Ordenação Externa



- Considere um exemplo de um arquivo de tamanho muito grande, tal como 1 bilhão de palavras. Considere ainda uma memória interna disponível de 2 milhões de palavras e que haja 8 unidades de fitas magnéticas para realização de uma ordenação por intercalação-de-4-caminhos. Neste caso $P(n) = 5$ e o número total de passadas, incluindo a primeira passada para obter os n/m blocos ordenados, é 6.
- Uma estimativa do tempo total gasto para ordenar este arquivo pode ser obtido multiplicando-se por 6 o tempo gasto para transferir o arquivo de uma fita para outra.



Algoritmos de ordenação – Ordenação Externa



- Para uma intercalação-de- f -caminhos foram utilizadas $2 \times f$ fitas nos exemplos mostrados. Para usar apenas $f + 1$ fitas basta encaminhar todos os blocos para uma única fita e, com mais uma passada, redistribuir estes blocos entre as fitas de onde eles foram lidos.
- No caso do exemplo de 22 registros, apenas 4 fitas seriam suficientes: a intercalação dos blocos a partir das fitas 1, 2 e 3 seria toda dirigida para a fita 4; ao final, o segundo e o terceiro blocos ordenados de 9 registros seriam transferidos de volta para as fitas 1 e 2; e assim por diante. O custo envolvido é uma passada a mais em cada intercalação.



Exercícios



1) Considere um arquivo armazenado em uma fita de entrada composto pelos registros:

45 78 66 99 24 57 5 15 1 5 28 10 81 8 34

Suponha que cada registro seja equivalente a uma palavra (ex.: palavra de 32 bits). Assuma ainda que a memória interna tenha espaço só para 3 palavras e o número de unidades de fitas magnéticas disponíveis é quatro. Mostre a ordenação do arquivo utilizando intercalação-de-2-caminhos.

2) Qual o número de passadas necessárias, incluindo a da primeira etapa, para ordenar um arquivo com um trilhão de palavras, onde a memória interna tenha capacidade para 100 milhões de palavras e 10 unidades de fitas magnéticas estejam disponíveis para uma ordenação por intercalação-de-5-caminhos?

3) Para a mesma situação mostrada no exemplo 1, mostre a ordenação por intercalação-de-3-caminhos, mas utilizando somente 4 fitas ao invés de 6.

