



UNIVERSIDADE FEDERAL DE OURO PRETO
Campus Morro do Cruzeiro

ANÁLISE EXPERIMENTAL DE COMPLEXIDADE DE ALGORITMOS APLICADOS À ORDENAÇÃO EXTERNA

Ana Laura Siqueira Mendes - 20.2.4091

Diego Demétrio Santos Rodrigues - 20.2.4093

Eder Bragança Pereira - 20.2.4011

Mateus do Carmo Eleto Hamade - 20.2.4186

Selio Guilherme da Silva Filho - 20.2.4107

Sumário

ANÁLISE EXPERIMENTAL DE COMPLEXIDADE DE ALGORITMOS APLICADOS À ORDENAÇÃO EXTERNA	1
1. INTRODUÇÃO	3
2. INTERCALAÇÃO BALANCEADA DE VÁRIOS CAMINHOS “2F FITAS”	4
ESPECIFICAÇÃO DOS TESTES	4
2.1 ARQUIVOS ORDENADOS ASCENDENTEMENTE	4
2.2 ARQUIVOS ORDENADOS DESCENDENTEMENTE	5
2.3 ARQUIVOS ORDENADOS ALEATORIAMENTE	6
3. INTERCALAÇÃO BALANCEADA DE VÁRIOS CAMINHOS “F+1 FITAS”	7
ESPECIFICAÇÃO DOS TESTES:	7
3.1 ARQUIVOS ORDENADOS ASCENDENTEMENTE	7
3.2 ARQUIVOS ORDENADOS DESCENDENTEMENTE	8
3.3 ARQUIVOS ORDENADOS ALEATORIAMENTE	9
4. QUICKSORT EXTERNO	10
ESPECIFICAÇÃO DOS TESTES	10
4.1 ARQUIVOS ORDENADOS ASCENDENTEMENTE	10
4.2 ARQUIVOS ORDENADOS DESCENDENTEMENTE	12
4.3 ARQUIVOS ORDENADOS ALEATORIAMENTE	13
5. CONSIDERAÇÕES FINAIS	14

1. INTRODUÇÃO

Os experimentos relatados no documento têm como principal objetivo um estudo mais profundo da complexidade de desempenho de metodologias de ordenação externa sobre os algoritmos de “Intercalação balanceada de vários caminhos 2F fitas”, “Intercalação balanceada de vários caminhos $F + 1$ fitas” e do “QuickSort” adequado a arquivo externo.

Para tanto, cada método foi devidamente implementado utilizando a linguagem de programação “C” e posteriormente testado, considerando arquivos de 100, 200, 2.000, 20.000, 200.000 e 471.705 registros ordenados ascendentemente, descendentemente e aleatoriamente.

Em cada teste, submete-se um arquivo que corresponde a uma base de dados de notas de 471.705 alunos do Ensino Superior que fizeram o “Provão” em 2003. Os registros contidos nesse arquivo serão ordenados ascendentemente considerando o campo “nota”. A análise de complexidade considerará os seguintes parâmetros: o número de transferências de itens da memória externa para a memória interna (leitura), número de transferências da memória interna para a memória externa (escrita), as comparações entre as chaves de ordenação e o tempo de execução da função correspondente ao método a ser testado.

Por fim, elaborar-se-á a comparativa de desempenho entre os métodos citados, a fim de inferir informações sobre eles.

2. INTERCALAÇÃO BALANCEADA DE VÁRIOS CAMINHOS “2F FITAS”

A intercalação balanceada de vários caminhos “2f fitas” foi implementada com base em uma memória interna disponível para armazenar um vetor de, no máximo, 10 registros e 20 fitas de armazenamento externo. Esse método utiliza metade das fitas totais como entrada e a outra metade como fitas de saída. As fitas de entrada são preenchidas de acordo com o funcionamento do Heap e posteriormente introduzidas na fita de saída, gerando blocos ordenados, que serão novamente redistribuídos nas fitas de entrada até estarem totalmente ordenados.

ESPECIFICAÇÃO DOS TESTES

2.1 ARQUIVOS ORDENADOS ASCENDENTEMENTE

ARQUIVO COM 100 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.000000 seg	101	100	500

ARQUIVO COM 200 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.000000 seg	201	200	1.000

ARQUIVO COM 2.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.001000 seg	2.001	2.000	10.000

ARQUIVO COM 20.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.006000 seg	20.001	20.000	100.000

ARQUIVO COM 200.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.064000 seg	200.001	200.000	1.001.386

ARQUIVO COM 471.705 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.155000 seg	471.706	471.705	2.362.914

2.2 ARQUIVOS ORDENADOS DESCENDENTEMENTE

ARQUIVO COM 100 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.000000 seg	108	100	727

ARQUIVO COM 200 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.003000 seg	416	400	1.615

ARQUIVO COM 2.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.005000 seg	4.063	4.000	20.949

ARQUIVO COM 20.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.027000 seg	60.161	60.000	239.519

ARQUIVO COM 200.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.228000 seg	600.427	600.000	2.726.539

ARQUIVO COM 471.705 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.526000 seg	1.415.861	1.415.118	7.031.292

2.3 ARQUIVOS ORDENADOS ALEATORIAMENTE

ARQUIVO COM 100 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.000000 seg	101	100	500

ARQUIVO COM 200 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.000000 seg	201	200	1.000

ARQUIVO COM 2.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.005000 seg	4.053	4.000	19.933

ARQUIVO COM 20.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.035000 seg	61.009	60.000	468.950

ARQUIVO COM 200.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.371000 seg	810.725	800.000	5.955.762

ARQUIVO COM 471.705 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
1.028000 seg	2.383.679	2.358.525	14.530.527

3. INTERCALAÇÃO BALANCEADA DE VÁRIOS CAMINHOS “F+1 FITAS”

A intercalação balanceada de F+1 fitas possui o mesmo modo de funcionamento do método de intercalação com 2F fitas, porém, utiliza-se, nessa implementação, de **19** fitas de entrada e apenas 1 fita de saída. Neste método as fitas de entrada são preenchidas com o elemento de menor valor. Para tanto, aplica-se a estrutura de um Min Heap, seleciona-se a menor posição e posteriormente seu valor é inserido na fita de saída, gerando blocos ordenados, que serão novamente redistribuídos nas fitas de entrada até estarem totalmente ordenados.

ESPECIFICAÇÃO DOS TESTES:

3.1 ARQUIVOS ORDENADOS ASCENDENTEMENTE

ARQUIVO COM 100 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.002000 seg	239	202	828

ARQUIVO COM 200 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.001000 seg	439	402	1.728

ARQUIVO COM 2.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.005000 seg	4.039	4.002	17.928

ARQUIVO COM 20.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.022000 seg	40.039	40.002	179.928

ARQUIVO COM 200.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.227000 seg	400.039	400.002	1.805561

ARQUIVO COM 471.705 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.508000 seg	943.449	943.412	4.263.044

3.2 ARQUIVOS ORDENADOS DESCENDENTEMENTE

ARQUIVO COM 100 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.001000 seg	239	206	1.044

ARQUIVO COM 200 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.001000 seg	439	409	2.496

ARQUIVO COM 2.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.017000 seg	8.118	8.048	34.336

ARQUIVO COM 20.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.078000 seg	80.198	80.130	380768

ARQUIVO COM 200.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.458000 seg	800.418	800.370	4.910.506

ARQUIVO COM 471.705 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
1.933500 seg	2.830.987	2.830.903	11.203.696

3.3 ARQUIVOS ORDENADOS ALEATORIAMENTE

ARQUIVO COM 100 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.000000 seg	239	204	1.206

ARQUIVO COM 200 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.001000 seg	439	407	3.077

ARQUIVO COM 2.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.021000 seg	8.118	8.059	50.295

ARQUIVO COM 20.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0.092000 seg	120.657	120.567	698.839

ARQUIVO COM 200.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
1.04500 seg	1.205.757	1.205.671	8.968.033

ARQUIVO COM 471.705 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
2.690000 seg	3.787.016	3.786.896	22.160.037

4. QUICKSORT EXTERNO

O método QuickSort adequado para ordenação externa, assim como o aplicado à ordenação interna, utiliza a metodologia de Divisão e Conquista. Ele é um dos mais utilizados por possuir complexidade logarítmica, $O(\log n)$, exceto no pior caso, onde a complexidade é $O(n^2)$. Esse método, diferentemente dos anteriores, realiza a ordenação *in situ*, não sendo necessário criar outros arquivos para realizar a ordenação.

Para esse teste, sua implementação considera a existência de memória interna disponível para armazenar um vetor de, no máximo, 10 registros.

ESPECIFICAÇÃO DOS TESTES

4.1 ARQUIVOS ORDENADOS ASCENDENTEMENTE

ARQUIVO COM 100 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0 seg	100	100	335

ARQUIVO COM 200 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0 seg	200	200	685

ARQUIVO COM 2.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,005 seg	2.000	2.000	6.985

ARQUIVO COM 20.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,047 seg	20.000	20.000	69.985

ARQUIVO COM 200.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,929 seg	397.781	397.781	1.745.563

ARQUIVO COM 471.705 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
2,204 seg	951.013	951.013	4.404.866

4.2 ARQUIVOS ORDENADOS DESCENDENTEMENTE

ARQUIVO COM 100 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0 seg	149	149	877

ARQUIVO COM 200 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0 seg	321	321	1.872

ARQUIVO COM 2.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,01 seg	3.783	3.783	20.517

ARQUIVO COM 20.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,092 seg	39.412	39.412	188.690

ARQUIVO COM 200.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,919 seg	399.711	399.711	1.874.701

ARQUIVO COM 471.705 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
2,207 seg	951.013	951.013	4.415.394

4.3 ARQUIVOS ORDENADOS ALEATORIAMENTE

ARQUIVO COM 100 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,001 seg	292	292	1.550

ARQUIVO COM 200 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,001 seg	719	719	3.229

ARQUIVO COM 2.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,03 seg	12.454	12.454	38.091

ARQUIVO COM 20.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
0,768 seg	253.245	253.245	464.810

ARQUIVO COM 200.000 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
53,879 seg	13.881.264	13.881.264	15.851.804

ARQUIVO COM 471.705 REGISTROS			
TEMPO DE EXECUÇÃO	Nº DE TRANSFERÊNCIAS (LEITURA)	Nº DE TRANSFERÊNCIAS (ESCRITA)	Nº DE COMPARAÇÕES
217,694 seg	54.288.777	54.288.777	58.959.641

5. CONSIDERAÇÕES FINAIS

Findados os testes, pudemos observar que, de modo geral, ao empregar a ordenação sobre arquivos previamente ordenados ascendentemente ou descendemente, o tempo de execução das funções, bem como a quantidade de comparações, diminui consideravelmente.

O método de pesquisa externa utilizando o Intercalação Balanceada de Vários Caminhos “2F fitas” e “F+1 fitas” obtiveram resultados semelhantes de desempenho quando analisa-se o tempo de execução. Ao verificar as comparações, no entanto, nota-se que o método “F+1 ” fitas tem um custo maior, pois necessita de uma passada a mais em cada intercalação.

O método QuickSort, por sua vez, apesar de ter sua eficiência comprovada, é um algoritmo que, ao ser aplicado a arquivos grandes, o tempo de execução sobe significativamente, podendo demorar minutos para finalizar sua execução. Esse fato é resultado da ordenação *in situ*, que aumenta a quantidade de leituras e escritas, funcionalidade essa que é muito custosa ao programa.

Por fim, cabe concluir que a escolha de um método de ordenação externa deve ser feito com base no tamanho e no tipo de ordenação inicial do arquivo. O QuickSort, embora seja um dos métodos mais utilizados, não é o mais eficiente em casos maiores, porém, é um método extremamente vantajoso, quando não há disponibilidade de memória externa, pois toda a ordenação é feita dentro do próprio arquivo. A Intercalação Balanceada de Vários Caminhos “2F fitas” e “F+1 fitas”, no que lhe diz respeito, é pouco eficiente em casos pequenos, exigindo muito mais comparações em relação ao QuickSort, mas mostrou-se efetivo para reduzir o tempo de execução ao ordenar arquivos com muitos registros.