



Trabalho Prático I – Pesquisa Externa

Grupo de 4 a 6 alunos - Valor: 15,0 pontos - Data de entrega: 23/11/23

O objetivo deste trabalho prático consiste em um estudo da complexidade de desempenho dos seguintes métodos de pesquisa externa apresentados em sala de aula: (1) acesso sequencial indexado, (2) árvore binária de pesquisa adequada à memória externa, (3) árvore B e (4) árvore B*.

A 1ª fase deste trabalho corresponde à implementação em C dos métodos mencionados, considerando arquivos binários de registros quaisquer e memória interna disponível para armazenar os índices necessários, quando for o caso. Os registros devem estar armazenados apenas nos arquivos e possuem os seguintes campos: chave (valor inteiro), dado1 (valor inteiro longo) e dado2 (cadeia de 5000 caracteres). No caso, o campo "chave" representa a chave de pesquisa, por meio da qual os índices necessários em memória principal devem ser construídos.

A 2ª fase deste trabalho corresponde à análise experimental da complexidade de desempenho dos métodos mencionados, considerando as etapas de criação dos índices necessários, quando for o caso, e da própria pesquisa, por meio dos seguintes quesitos:

- número de transferências (leitura) de itens da memória externa para a memória interna;
- número de comparações entre chaves de pesquisa;
- tempo de execução (tempo do término de execução menos o tempo do início de execução).

Deve-se analisar, **separadamente**, as etapas de criação dos índices e da própria pesquisa quanto aos quesitos definidos.

Para a 2ª fase, experimentos devem ser realizados considerando arquivos contendo 100, 200, 2.000, 20.000, 200.000 e 2.000.000 registros quaisquer cujas chaves serão pesquisadas. Ademais, as seguintes distintas situações de ordem de arquivo devem ser consideradas para cada conjunto de registros: arquivos ordenados ascendentemente pela chave de pesquisa, arquivos ordenados descendentemente pela chave de pesquisa, e arquivos desordenados aleatoriamente pela chave de pesquisa (é importante ressaltar que, de acordo com o método de pesquisa externa a ser executado, algumas situações de ordem de arquivo não se aplicam). Para cada experimento, considerando os mesmos parâmetros (método de ordenação, quantidade de registros, situação de ordem do arquivo), deve-se ocorrer a pesquisa automática de 20 chaves de pesquisa distintas, bem diferenciadas e existentes no arquivo em questão, no intuito de se obter a média de cada um dos quesitos a serem considerados no processo de análise experimental.

Independente dos testes a serem efetuados, o programa **deve** ser implementado de tal forma que seja possível executá-lo, livremente, a partir da seguinte linha de comando no console:

pesquisa <método> <quantidade> <situação> <chave> [-P]

onde:

- <método> representa o método de pesquisa externa a ser executado, podendo ser um número inteiro de 1 a 4, de acordo com a ordem dos métodos mencionados;
- <quantidade> representa a quantidade de registros do arquivo considerado;
- <situação> representa a situação de ordem do arquivo, podendo ser 1 (arquivo ordenado ascendentemente), 2 (arquivo ordenado descendentemente) ou 3 (arquivo desordenado aleatoriamente);
- <chave> representa a chave a ser pesquisada no arquivo considerado;
- [-P] representa um argumento opcional que deve ser colocado quando se deseja que as chaves de pesquisa dos registros do arquivo considerado sejam apresentadas na tela.

Uma execução do comando "pesquisa" deve retornar se a chave de pesquisa desejada foi ou não encontrada no arquivo. Uma vez tendo sido encontrada, o registro completo da chave de pesquisa e os



valores referentes aos quesitos de análise (o número de transferências (leitura) entre as memórias interna e externa, o número de comparações entre as chaves de pesquisa e o tempo de execução) devem ser também retornados pelo comando. Observe que, no momento da execução do comando, o arquivo binário de registros deve já existir ou ser criado automaticamente de acordo com os valores especificados para os argumentos <quantidade> e <situação>. O tempo gasto para a criação de tal arquivo **não** faz parte do tempo de execução do método, um dos quesitos de análise experimental.

Para fins de avaliação deste trabalho, devem ser entregues:

1. Código-fonte do programa, bem indentado e comentado.
2. Relatório contendo os seguintes pontos:
 - Introdução: especificação do problema a ser resolvido; descrição dos testes realizados.
 - Para cada método: análise dos quesitos especificados considerando todos os testes realizados sobre o método.
 - Conclusão: análise comparativa de desempenho entre os métodos; principais dificuldades de implementação dos métodos.

Observações Importantes:

- Toda mensagem de orientação e de erro deve ser devidamente tratada.
- O código-fonte do programa deve estar **bem indentado e comentado**.
- Trabalhos copiados terão suas notas divididas pelo número de cópias.
- Penalização por atraso: 2,5 pontos a cada aula.
- Para não ocorrer perda na pontuação do trabalho por atraso:
 - o código-fonte e o relatório solicitado devem ser submetidos ao Moodle até às 10h do dia 23/11;
 - o programa deve ser apresentado pelo grupo ao professor na aula do dia 23/11.