

Projeto 02

Disciplina: *Organização de Arquivos – Turma C*

Semestre: 2º/2017

Prof.: Flávio Vidal

Título: *Geração de Índices com registros de tamanho variável utilizando árvores B (B-tree)*

Entrega: 15/12/2017 – 23h *Relatório e Arquivos-Fontes via Aprender.unb.br*

1. Objetivos

Implementar, utilizando obrigatoriamente a linguagem C/C++ ou Python procedimentos de indexação de arquivos (geração de índices primários) com registros com campos de tamanho variável, utilizando técnicas de árvores B (B-trees) para a geração dos arquivos de índices de forma a aplicar os conceitos básicos de técnicas de indexação. Este projeto é parte integrante da ementa da disciplina de Organização de Arquivos.

O projeto desta disciplina é uma atividade planejada de forma a complementar e reforçar o conteúdo programático da disciplina Organização de Arquivos. Espera-se que nas atividades de projeto os alunos desenvolvam sua capacidade de observação, análise e compreensão das metodologias de organização e estruturas de arquivos.

Desta forma cabe ao aluno, partindo da premissa que possui os requisitos para o curso, juntamente com o conteúdo adquirido nas aulas teóricas, desenvolver todas as etapas da implementação solicitada.

2. Introdução Teórica

Todos os aspectos teóricos podem ser encontrados no Capítulo 8 (incluindo anexos) do livro texto, complementando com as transparências apresentadas em sala de aula sobre o assunto.

3. Metodologia

Deverá ser implementado utilizando linguagem de programação de alto nível (linguagem C/C++ ou Python para Linux/Unix de preferência) uma aplicação que atenda os seguintes requisitos:

Requisito 1: Seja capaz de realizar a leitura de arquivos do benchmark do arquivo data.txt (disponível no ambiente aprender.unb.br da disciplina) com os campos de tamanho variável utilizando “,” (vírgula) como caractere separador de campos. Neste caso em específico o arquivo fornecido é um *dumping* de um banco de dados público na *web*¹ em que o primeiro campo de cada registro é a chave primária.

¹ Disponível em http://spatialkeydocs.s3.amazonaws.com/FL_insurance_sample.csv.zip.

Requisito 2: Criar arquivo de índice e de dados

A partir do arquivo de benchmark disponibilizado, crie um arquivo de índices primário à parte construído utilizando árvores B (*indicelista.bt*), em que o elemento (conteúdo) do nó seja a chave primária do arquivo base (primeiro campo de cada registro). Observe que antes da criação da árvore B deve-se verificar a integridade do arquivo de entrada, principalmente referente às repetições de registros com a mesma chave primária.

Crie uma rotina para a visualização do arquivo de índice na forma de árvore B no monitor. Durante o processo de criação da árvore B o fator de ramificação e sua ordem deverão ser definidas dinamicamente a partir da quantidade de registros existentes no arquivo de benchmark de entrada (Vide aulas de árvores B para maiores detalhes).

Requisito 3: Regravar arquivo de índice após o uso.

Defina um conjunto de rotinas básicas necessárias para após a utilização do arquivo de índice, este seja regravado em disco após as operações sejam realizadas, tanto no arquivo de índice, quanto no arquivo de dados.

Requisito 5: Inclusão e Busca de registros

Deverá ser desenvolvido rotinas que permitam a inclusão de novos registros e busca no arquivo de dados a partir do processamento do arquivo de índices das árvores B gerado. Lembrando que o arquivo de índices deverá ser mantido na forma de árvore B. Mostrar no terminal de vídeo o arquivo de índice antes e depois da inclusão/busca e listar em monitor quais os nós foram percorridos até o final da busca e/ou inclusão de novos registros.

Requisito 6: Excluir registros

Deverá ser desenvolvido rotinas que permitam a exclusão de novos registros no arquivo de dados e também no arquivo de índices. Lembrando que o arquivo de índices deverá ser mantido seguindo rigorosamente os princípios de manutenção das árvores B e metodologias de re-utilização do espaço disponibilizado. Mostrar no terminal de vídeo o arquivo de índice antes e depois da exclusão.

Requisito 7: Atualizar registros:

Definir rotinas nas quais implementem a atualização de registros nos arquivos de dados, e respectivamente nos arquivos de índices, seguindo as condições descritas abaixo:

Requisito 7.1: Com mudança de chave primária.

Requisito 7.2: Sem mudança de chave primária.

Deve-se ressaltar que estas rotinas, implementadas em cada requisito, deverão contemplar todos os aspectos relativos às características dos arquivos de índices, de forma que todas as operações realizadas, tanto no arquivo de índices, quanto as realizadas nos arquivos de dados (arquivos de *benchmarks*), devem ser apresentadas via terminal de vídeo, e ainda preservando o arquivo gerado em disco.

4. Grupos

Neste projeto será permitido a formação de grupos com no máximo 3 (três) alunos. A partir do grupo formado, deverá ser indicado um líder, no qual este líder será o responsável pelo envio dos arquivos fontes para o sistema Aprender.unb.br.

Somente serão aceitos os arquivos fontes enviado pelo líder do grupo. O relatório escrito deverá ser entregue em PDF na área específica da atividade no ambiente aprender.unb.br. Reitero que não será aceito nenhum arquivo (e/ou relatório) via e-mail do professor, independente de indisponibilidade do ambiente e/ou outro problema decorrente.

5. Relatório

O relatório deve demonstrar que a respectiva atividade de laboratório foi realizada com sucesso e que os princípios subjacentes foram compreendidos.

O relatório da atividade de laboratório é o documento gerado a partir do trabalho realizado seguindo as orientações exigidas na metodologia de laboratório. Este deve espelhar todo o trabalho desenvolvido nos processos de obtenção dos dados e sua análise. Apresentamos a seguir uma recomendação de organização para o relatório da atividade de laboratório. Deverá conter as seguintes partes:

i. Identificação: Possuir a indicação clara do título do experimento abordado, a data da sua realização, a identificação da disciplina/turma, os nomes dos componentes do grupo, número de matrícula e e-mail.

ii. Objetivos: Apresentar de forma clara, porém sucinta, os objetivos do laboratório.

iii. Introdução: Deve conter a teoria necessária à realização da atividade de laboratório.

iv. Materiais e Métodos: É dedicada à apresentação dos materiais e equipamentos, descrição do arranjo experimental e uma exposição minuciosa do procedimento de laboratório realmente adotado.

v. Resultados: Nesta parte são apresentados os resultados das implementações efetuadas, na forma de tabelas e gráficos, sem que se esqueça de identificar em cada caso os parâmetros utilizados.

vi. Discussão e Conclusões: A discussão visa comparar os resultados obtidos e os previstos pela teoria. Deve se justificar eventuais discrepâncias observadas. As conclusões resumem a atividade de laboratório e destacam os principais resultados e aplicações dos conceitos vistos.

vii. Bibliografia: Citar as fontes consultadas, respeitando as regras de apresentação de bibliografia (autor, título, editora, edição, ano, página de início e fim).

O relatório do laboratório deverá ser confeccionado em editor eletrônico de textos, utilizando o padrão de formatação descrito no arquivo de exemplo, disponibilizado no website da disciplina (oficial e/ou mirror). Está disponibilizado um único padrão de formatação para editores científicos LATEX (arquivo extensão *.zip contendo arquivo de exemplo do uso do pacote), cabendo ao grupo a escolha de qual editor Latex será utilizado. Este modelo pode ser acessado no aprender da disciplina. Somente serão aceitos para avaliação relatórios em PDF feitos seguindo esta formatação.

Todo o código fonte (inclusive as diretrizes de compilação utilizadas) deverá ser entregue via *upload* no ambiente Aprender.unb.br, em arquivo *.zip completando a atividade designada ao laboratório correspondente (vide ambiente aprender.unb.br para maiores detalhes). O código fonte deverá ser “re-compilável” para que seja realizada a correção no ambiente computacional Linux. Não serão aceitos trabalhos entregues fora do prazo estipulado, sendo atribuída nota zero ao grupo. Não serão

aceitos qualquer tipo de material (relatório e códigos-fonte) via e-mail do professor. O único método de envio deverá ser feito pelo *Aprender.unb.br*.

Vale ressaltar que será atribuída nota zero, definida como atividade “incompleta”, ao grupo que não entregar o relatório e/ou código fonte implementado e devidamente identificado. Entende-se como atividade completa versão em PDF do relatório e arquivos-fonte corretamente enviados ao endereço eletrônico *Aprender.unb.br*.

a. Critérios Empregados na Correção do Relatório de Laboratório

A avaliação dos relatórios terá em consideração os seguintes itens:

<i>No.</i>	<i>Item</i>	<i>Descrição</i>	<i>Peso (%)</i>
1	Apresentação	Qualidade dos gráficos, impressão, tabelas, vocabulário, legendas, etc.	10%
2	Aspectos Teóricos	Apresentação e descrição da base teórica utilizada. Avaliação da bibliografia utilizada se necessário.	20%
3	Materiais e Métodos	Descrição de todos os procedimentos utilizados, contemplando dados técnicos, bem como a metodologia utilizada no decorrer do projeto.	10%
4	Resultados	Todos os resultados alcançados no projeto.	30%
5	Discussão e Conclusões	Discussão objetiva e devidamente explicada a respeito do projeto. Inclui-se também a pontuação por iniciativa.	30%

Dúvidas deverão ser encaminhadas ao fórum de discussão específico no ambiente *Aprender.unb.br*.