

Trabalho Prático 04

Henrique Oliveira, Gabriel Jota

¹Instituto de Ciências Exatas e Informática – Pontifícia Universidade Católica de Minas Gerais (PUCMG)

Abstract. *This paper presents the development and implementation of advanced algorithms and data structures as part of the course "Algorithms and Data Structures 3" in the Computer Science program. The project was divided into four practical assignments, each focusing on different techniques: sequential file manipulation, indexed file handling using B+ Trees, Hashing and Inverted Lists, data compression with Huffman and LZW, and pattern matching and cryptography. The Zara clothing products dataset was used to test and validate the implementations. Challenges included adapting algorithms to the dataset format, optimizing performance, and ensuring data integrity. Solutions involved modular implementation, preprocessing of data, and automated testing. This report details the development process, encountered challenges, and achieved results.*

Resumo. *Este artigo apresenta o desenvolvimento e implementação de algoritmos avançados e estruturas de dados como parte da disciplina "Algoritmos e Estruturas de Dados 3" do curso de Ciência da Computação. O projeto foi dividido em quatro trabalhos práticos, cada um focando em técnicas diferentes: manipulação de arquivos sequenciais, manipulação de arquivos indexados utilizando Árvores B+, Hashing e Listas Invertidas, compactação de dados com Huffman e LZW, e casamento de padrões e criptografia. A base de dados de produtos de roupa da Zara foi utilizada para testar e validar as implementações. Os desafios incluíram a adaptação dos algoritmos ao formato da base de dados, a otimização do desempenho e a garantia da integridade dos dados. As soluções envolveram a implementação modular, o pré-processamento dos dados e a realização de testes automatizados. Este relatório detalha o processo de desenvolvimento, os desafios enfrentados e os resultados obtidos.*

1. Introdução

No curso de Ciência da Computação, a disciplina "Algoritmos e Estruturas de Dados 3" proporcionou um estudo aprofundado de algoritmos avançados e técnicas de manipulação de dados. Durante este período, os alunos foram desafiados a aplicar esses conceitos em um projeto prático dividido em quatro etapas. Cada etapa abordou diferentes aspectos de algoritmos e estruturas de dados, culminando em um relatório final e um sistema integrado. Este relatório documenta o desenvolvimento e a implementação dos Trabalhos Práticos (TPs), detalhando os desafios enfrentados, as soluções encontradas e os resultados obtidos.

A base de dados utilizada para os trabalhos práticos foi a de produtos de roupa da Zara, disponível no Kaggle. A utilização dessa base de dados apresentou uma série de desafios práticos, desde a adaptação do código para compatibilidade com o formato

dos dados até a implementação de algoritmos eficientes para resolver problemas específicos. Este relatório detalha cada uma das quatro etapas, desde a concepção inicial até a realização dos testes e análise dos resultados.

2. Desenvolvimento

2.1. TP1: Criação da Base de Dados e Manipulação de Arquivo Sequencial

A primeira etapa envolveu a criação da base de dados a partir dos dados brutos da Zara e a implementação da manipulação de arquivos sequenciais. O objetivo foi organizar e armazenar os dados de forma eficiente, permitindo operações básicas de inserção, deleção e busca.

Desafios Enfrentados

- **Formatação dos Dados:** A criação da base de dados a partir dos arquivos brutos exigiu uma formatação cuidadosa dos dados, garantindo a consistência e integridade das informações.
- **Manipulação de Arquivos:** Implementar a manipulação de arquivos sequenciais de maneira eficiente e robusta foi um desafio, especialmente para garantir o desempenho adequado durante as operações de inserção e busca.

Soluções Implementadas

- **Scripts de Formatação:** Desenvolvemos scripts para automatizar a formatação e limpeza dos dados, facilitando a criação da base de dados.
- **Estrutura de Dados Otimizada:** Implementamos uma estrutura de dados sequencial otimizada para realizar operações de inserção, deleção e busca de maneira eficiente.

2.2. TP2: Manipulação de Arquivo Indexado com Árvore B, Hash e Lista Invertida

A segunda etapa focou na implementação de técnicas avançadas de indexação utilizando Árvore B, Hashing e Lista Invertida para melhorar a eficiência das operações de busca na base de dados.

Desafios Enfrentados

- **Complexidade das Estruturas de Indexação:** Implementar e otimizar as estruturas de Árvore B+, Hash e Lista Invertida exigiu um entendimento profundo de suas características e comportamento.
- **Integração com a Base de Dados:** Integrar essas estruturas de indexação com a base de dados existente sem comprometer a integridade e a performance foi um desafio significativo.

Soluções Implementadas

- **Implementação Modular:** Adotamos uma abordagem modular para implementar cada estrutura de indexação, facilitando a integração e manutenção do código.
- **Otimização de Índices:** Aplicamos técnicas de otimização para garantir que os índices melhorassem significativamente o desempenho das operações de busca sem sobrecarregar o sistema.

2.3. TP3: Compactação com Huffman e LZW

Na terceira etapa, implementamos algoritmos de compactação de dados, especificamente Huffman e LZW, para reduzir o espaço de armazenamento necessário para a base de dados.

Desafios Enfrentados

- Implementação dos Algoritmos de Compactação: A implementação correta dos algoritmos de Huffman e LZW foi complexa devido às suas características intrínsecas e à necessidade de garantir a compactação e descompactação corretas.
- Análise de Eficiência: Avaliar a eficiência dos algoritmos em termos de taxa de compressão e tempo de processamento foi um desafio adicional.

Soluções Implementadas

- Desenvolvimento Modular: Optamos por uma abordagem modular para implementar os algoritmos, facilitando a manutenção e a correção de erros.
- Ferramentas de Análise: Utilizamos ferramentas de análise para medir a taxa de compressão e o tempo de processamento, permitindo uma comparação detalhada entre os algoritmos.

2.4. TP4: Casamento de Padrões e Criptografia

A quarta e última etapa consistiu na implementação de algoritmos de casamento de padrões e criptografia, sendo estes Boyer-Moore e RSA.

Desafios Enfrentados

- Complexidade dos Algoritmos: A implementação correta dos algoritmos de casamento de padrões e criptografia exigiu um entendimento profundo de seus funcionamentos e otimizações.
- Integração dos Algoritmos: Integrar os diferentes algoritmos de casamento de padrões e criptografia com o sistema existente apresentou desafios de coordenação e compatibilidade.

Soluções Implementadas

- Arquitetura Modular: Desenvolvemos uma arquitetura modular que permitiu a integração suave dos diferentes algoritmos.

3. Testes e Resultados

Ao longo de cada um dos trabalhos práticos, com exceção do TP04, foi feito um vídeo que explica a utilização do sistema, que por sua vez inclui casos de teste de cada um dos algoritmos, bem como a inserção, atualização, leitura e exclusão de registros da base de dados.

4. Conclusão

A realização dos quatro Trabalhos Práticos ao longo da disciplina "Algoritmos e Estruturas de Dados 3" proporcionou uma experiência rica e abrangente, consolidando o entendimento teórico e prático de diversos algoritmos e estruturas de dados. Cada etapa do projeto contribuiu significativamente para a compreensão de diferentes aspectos de sistemas de gerenciamento de banco de dados (SGBD) e algoritmos avançados.

Na primeira etapa, a criação da base de dados e a manipulação de arquivos sequenciais permitiram uma imersão nas operações básicas de um SGBD, como inserção, deleção e busca. A necessidade de garantir a integridade e consistência dos dados, além de otimizar o desempenho das operações, reforçou a importância das boas práticas no gerenciamento de dados e no desenvolvimento de sistemas robustos.

A segunda etapa, que envolveu a manipulação de arquivos indexados com Árvores B+, Hashing e Listas Invertidas, expandiu o entendimento sobre técnicas avançadas de indexação. A implementação dessas estruturas evidenciou como elas podem melhorar significativamente a eficiência das operações de busca, mostrando na prática o impacto direto das estruturas de dados na performance de um SGBD.

Na terceira etapa, os algoritmos de compactação Huffman e LZW foram implementados, destacando a importância da compactação de dados para a economia de espaço de armazenamento e eficiência na transmissão de dados. A prática de desenvolver esses algoritmos e avaliar suas taxas de compressão e tempos de processamento proporcionou uma visão detalhada dos trade-offs envolvidos na escolha de técnicas de compactação.

Finalmente, na quarta etapa, a integração de algoritmos de casamento de padrões e criptografia, incluindo Boyer-Moore e RSA, completou o ciclo de aprendizado. A implementação desses algoritmos evidenciou a aplicação prática das técnicas de busca e segurança de dados, essenciais para o desenvolvimento de sistemas que lidam com grandes volumes de informações e que exigem proteção contra acessos não autorizados.

Os desafios enfrentados ao longo do desenvolvimento dos trabalhos práticos, desde a adaptação dos algoritmos ao formato da base de dados até a otimização do desempenho e a integração dos diferentes módulos, proporcionaram uma experiência de aprendizado inestimável. As soluções encontradas, como a implementação modular, o pré-processamento de dados e a realização de testes automatizados, demonstraram a importância de uma abordagem estruturada e metodológica no desenvolvimento de sistemas complexos.

Em suma, os Trabalhos Práticos não só reforçaram o conhecimento teórico sobre algoritmos e estruturas de dados, mas também proporcionaram uma compreensão profunda de sua aplicação prática em sistemas reais. Essa experiência preparou de maneira eficaz para enfrentar desafios futuros no campo da Ciência da Computação, equipando com habilidades e conhecimentos essenciais para o desenvolvimento e gerenciamento de sistemas de dados eficientes e seguros.