



Disciplina Integradora II

OTIMIZAÇÃO POR LEVENSHTEIN

Eduardo C. Andrade – Marcelo P. R. Heredia – Michael L. S. Rosa

Orientador: Duncan Dubugras Alcoba Ruiz

Introdução



OTIMIZAR TEMPO DE
CONSULTA A BASE DE DADOS



INDEXAÇÃO DE TABELAS

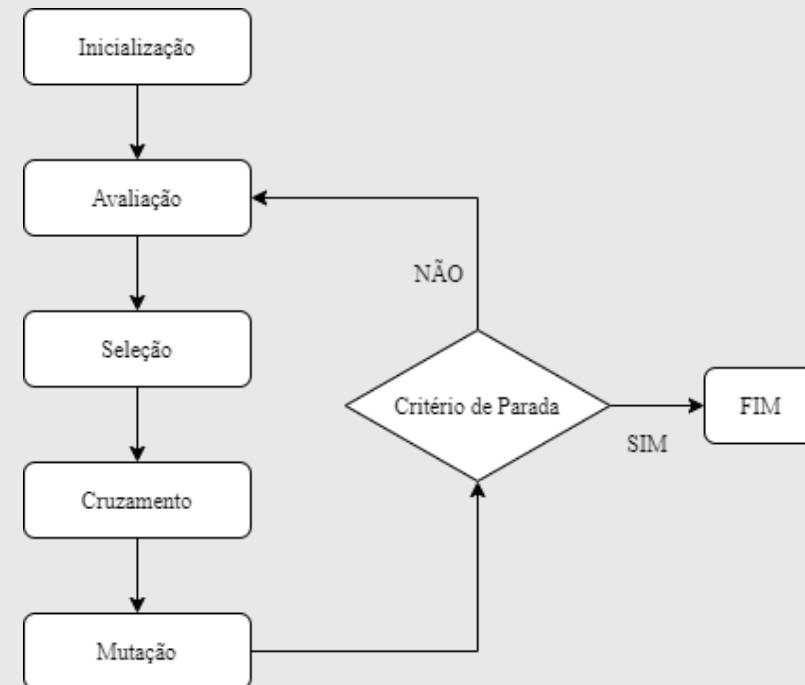


ALGORITMO GENÉTICO

Fundamentação Teórica

Algoritmo Genético

- Algoritmo baseado na teoria da evolução de Charles Darwin
- Busca as melhores soluções para problemas de busca e otimizações

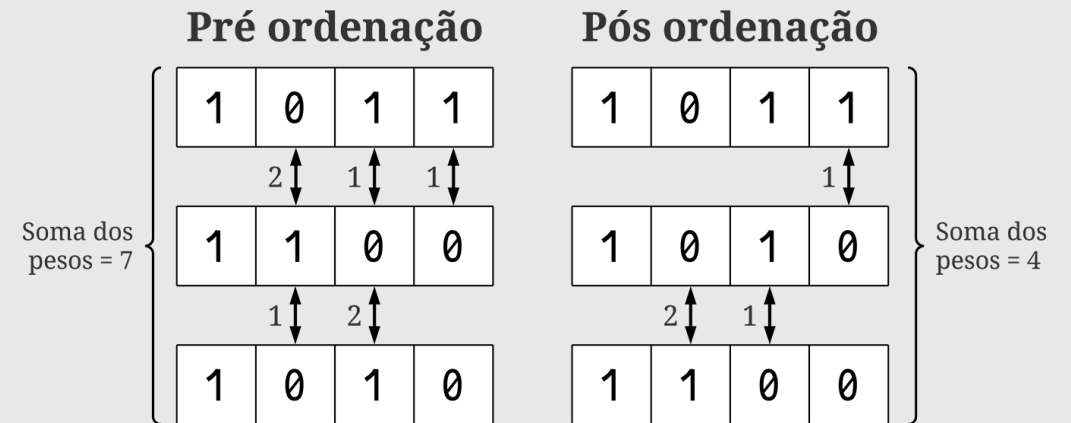


Fonte: Elaborado pelos autores

Fundamentação Teórica

Algoritmo de Levenshtein

- Distância de edição de palavras.
- Operações: inserção, remoção e substituição de caracteres.
- Possibilidade de aplicar pesos para operações e caracteres específicos.



Fonte: Elaborado pelos autores

Desenvolvimento



CRIAÇÃO DA BASE DE DADOS
USANDO TPC-H 500MB



DESENVOLVIMENTO DA
ORDENAÇÃO POR LEVENSHTTEIN

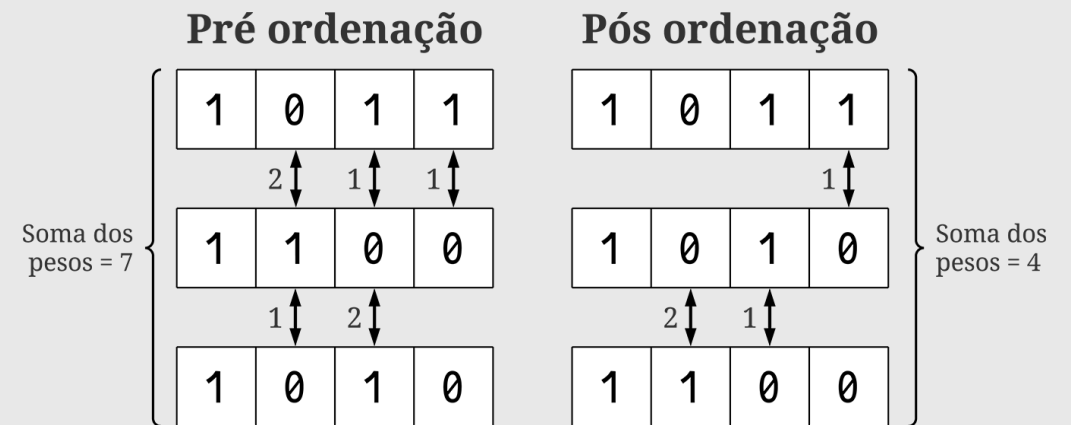


ADAPTAÇÃO E CORREÇÃO DO
ALGORITMO GENÉTICO

Desenvolvimento

Algoritmo de Levenshtein

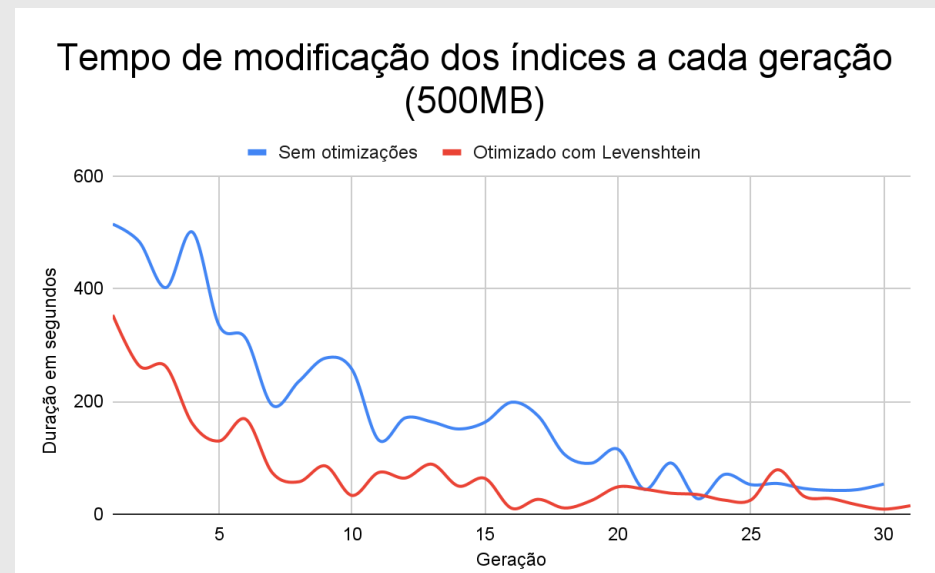
- Apenas operação de substituição e alfabeto de 2 caracteres ('1' ou '0').
- Início da ordenação baseia-se no estado atual do banco de dados.
- Substituições onde ocorre a criação a do índice pesam mais.
- Ordenação baseia-se apenas nas colunas com mais dados.



Fonte: Elaborado pelos autores

Resultados

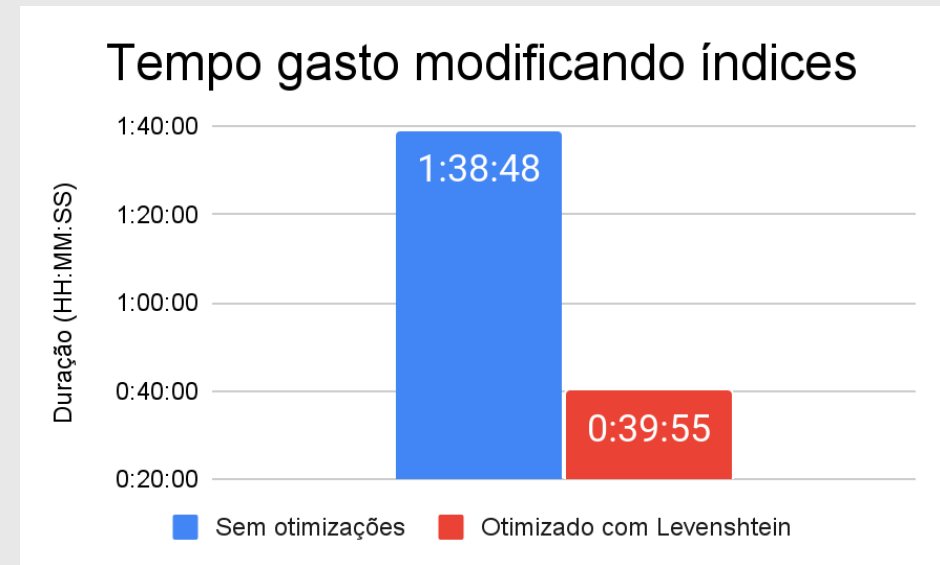
- O tempo gasto modificando os índices foi notavelmente menor após a otimização.
- A convergência da população faz com que diminuam as modificações nos índices.



Fonte: Elaborado pelos autores

Resultados

- O tempo total gasto modificando os índices diminuiu em quase uma hora após a otimização.
- A redução foi de quase 60% em relação ao não otimizado.



Fonte: Elaborado pelos autores

Trabalhos Futuros

Pesos por coluna

- Atribuir diferentes pesos para as colunas dependendo do tamanho.
- Deve fazer com que índices que demoram mais para ser criados pesem mais no cálculo de distância.

Funcionamento do SGBD

- Checar quais índices estão sendo utilizados.
- Entender otimizações que estão sendo feitas pelo SGBD e como influenciam o tempo de execução.

Trabalhos Relacionados

SmartIX: A database indexing agent based on reinforcement learning

- Seleção de índices através de aprendizado por reforço.

Genetic Algorithm for Database Indexing

- Esquemas de consultas e dados próprios.

Obrigado!

Contato:

- eduardo.a@edu.pucrs.br
- marcelo.heredia@edu.pucrs.br
- michael.rosa@edu.pucrs.br

Repositório com a otimização:

- <https://github.com/tecnicasilegais/Integradora2>