**Centro de Ensino Unificado de Teresina - CEUT**

Dann Luciano de Menezes

**Algoritmos Genéticos na Otimização de Consultas por Similaridade**

Teresina, PI

2010

**Centro de Ensino Unificado de Teresina - CEUT**

Dann Luciano de Menezes

**Algoritmos Genéticos na Otimização de Consultas por Similaridade**

Projeto apresentado como requisito para conclusão da disciplina de Trabalho de Conclusão de Curso I e pré requisito para cursar a disciplina de Trabalho de Conclusão de Curso II

Teresina, PI

2010

**Sumario**

**1 Tema**

**2 Problema**

**3 Objetivos**

**4 Justificativa**

**5 Metodologia**

**6 Referenciais Teóricos**

**6.1 Consultas por Similaridade**

**6.2 Algoritmos Genéticos**

**7 Cronograma**

**8 Recursos**

**9 Referências**

1. **Tema**

Algoritmos Genéticos na Otimização de Consultas por Similaridade.

1. **Problema**

Como otimizar o tempo computacional na obtenção de consultas por similaridade? A técnica dos Algoritmos Genéticos otimizaria o processo?

1. **Objetivos**

* Implementar uma ferramenta de busca em Domínios Métricos utilizando técnicas dos Algoritmos Genéticos.
* Acelerar as operações de busca de dados em Domínios Métricos

1. **Justificativa**

O tempo computacional gasto para realizar a operação de recuperação de dados em SGBDs é relativamente baixa, considerando o volume de informações armazenadas, isso ocorre somente nos tipos de dados mais simples como números e letras. Quando aumentamos a complexidade do dado a ser trabalhado aumentamos também o tempo computacional do processo de busca. O pior caso é quando temos que trabalhar com dados binários, como imagens, musicas, vídeos. Esses dados complexos são armazenados em estruturas de dados especificas nomeadas de Domínios Métricos. Apesar das características de indexação dos Domínios Métricos as operações de consultas dos dados são muitas vezes ineficientes devido ao problema da “maldição da dimensionalidade” apud (BUENO Renato, TRAINA Agma J. M. e Traina Caetano Jr. ***~~Ano~~***). ***Buendo; Traina; Traina (Ano)*** ainda afirmam que “em alguns casos a busca sequencial mais vantajosa”. Neste trabalho, procuramos mostrar uma abordagem diferente e eficiente pois segundo Zuben (***~~ANO~~***) são os problemas de otimização que mais recebem contribuições a partir das técnicas de computação evolutiva. Sendo os Algoritmos Genéticos a técnica que iremos utilizar neste trabalho.

1. **Metodologia**

1. **Referenciais Teóricos**

**6.1 Consultas por Similaridade**

Os primeiros Sistemas Gerenciados de Base de Dados (SGBDs) tinham a finalidade de facilitar e agilizar o processo de recuperação exata dos dados. Essa recuperação exata foi e continua sendo uma das necessidades das aplicações que os SGBDs devem suportar. Com o passar do tempo novas necessidades foram surgindo, e a recuperação exata dos dados deixar de ser necessária em algumas aplicações especificas como, sistemas de apoio à decisão. Isso porque essas aplicações possuem uma característica exploratória segundo ***Buendo; Traina; Traina (Ano)***. Outra necessidade é o armazenamento e recuperação de tipos de dados complexos, como já foi afirmado acima. As operações de recuperação não são realizadas em cima dos dados complexos propriamente dito e sim de características extraídas quando os dados são inseridos ou alterados. Os dados complexos devem passar antes por extratores de características que codificam as características dos dados complexos em números e depois adicionam a codificação nos Domínios Métricos existentes, para depois as operações de busca utilizam os Domínios Métricos.

**6.1.1 Domínios Métricos**

Um domínio métrico é definido como M = (D, d( )), onde D é o conjunto de todos os objetos que atendem às propriedades do domínio e d( ) é uma função de distância, ou métrica, entre esses objetos. Dados x,y e z E D, uma função de distancia (métrica) d : D x D 🡪 R+ deve satisfazer as seguintes propriedades:

1. *Simetria*: d(x,y) = d(y,x);
2. *Não Negatividade*: 0 <= d(x,y) <= 8, d(x,x) = 0;
3. *Desigualdade Triangular*: d(x,y) <= d(x,z) + d(z,y);

Em domínios métricos, os principais operadores de consulta por similaridade são:

**RQ(sq, rq) – Consulta por abrangência (“Similarity Range Query”):** retorna todos os objetos cujas distâncias até objeto central da consulta sq sejam menores ou iguais ao raio da consulta rq;

**k-NNQ(sq, k) – Consulta aos k-vizinhos mais próximos (“k-Nearest Neighbor Query”):** retorna os k objetos mais próximos do objeto central de consulta sq;

**6.1.2 Métodos de Acesso Métrico**

**6.2 Algoritmos Genéticos**

1. **Cronograma**

1. **Recursos**

**Referências**

BUENO Renato, TRAINA Agma J. M. e Traina Caetano Jr. **Algoritmos Genéticos para Consultas por Similaridade**.

ZUBEN Fernando J. Von. **Computação Evolutiva: Uma Abordagem Pragmática**

GOLDBERG David E. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley Company Inc. 1989

RUSSEL Stuart e NORVIG Peter. Inteligência Artificial. 2˚ Ed. Rio de Janeiro: Editora Campus. 2004.