

Um Algoritmo Genético para Geração de Strings em Estudos Terciários

Marco A. S. Martins¹, Welington Jr. L de Godoy¹, Francisco Carlos Souza¹

¹Engenharia de Software – Universidade Tecnológica Federal do Paraná (UTFPR)
CEP 85660-000 – Dois Vizinhos – PR – Brazil

marcomartins.1997@aluno.utfpr.edu.br, welingtongodoy@alunos.utfpr.edu.br,
franciscosouza@utfpr.edu.br

Abstract. *This experiment aims to improve or search for articles with keywords, their synonyms, and related keywords or close to initial searches. Problem: Often it is not possible to find applicable articles or excerpts within word search articles, simply because these words cannot be present at the same time or often just written with words, which have the same meaning but with another way of writing, this way many articles or excerpts are not found and leave study tools. With that in mind, we use a genetic algorithm through Hill Climb to generate newer Search Strings at the time of the search.*

Resumo. *Este experimento tem por objetivo o melhoramento da busca de artigos através de palavras chaves, seus sinônimos e palavras relacionadas ou próximas a utilizada na pesquisa inicial. Problema: Muitas vezes não é possível encontrar artigos específicos ou trechos dentro de artigos através da busca por palavras específicas, simplesmente por estas palavras não estarem presentes nos mesmos ou muitas vezes estando simplesmente escritas com palavras sinônimos, que contém o mesmo significado mas com outra forma de se escrever, desta maneira muitos artigos ou trechos não são encontrados e deixam de ser ferramentas de estudos. Pensando nisso utilizamos um algoritmo genético para gerar Strings de Busca mais eficazes no momento da pesquisa.*

Keywords

String de Busca, Algoritmo Genético.

1. Introdução

Na busca por fontes de conhecimento para realização de pesquisas muitas vezes utiliza-se de artigos específicos da área em questão. Durante as pesquisas por estes artigos muitas opções são encontradas fazendo com que a abrangência da pesquisa se estenda a áreas em que não se tem a necessidade de utilizar o conhecimento neles aplicados, assim, perde-se um valioso tempo de pesquisa com a leitura destes artigos desnecessários para a situação. Pensando em situações como esta que este trabalho tem como objetivo a criação de uma String de Busca combinada a um algoritmo genético que realize uma melhora nos resultados obtidos em pesquisas de artigos de áreas de conhecimento ao se pesquisar palavras chaves e seus sinonimos presentes nos mesmos, assim diminuindo a margem de artigos pouco relacionados ao tema abordado, dando ênfase ao que realmente se faz necessário.

2. Background

O uso de Strings de Busca como forma de melhorar às pesquisas de artigos científicos em áreas de conhecimentos específicos tem se tornado popular pois a melhora no encontro de resultados satisfatórios têm se tornado amplamente agradável para aqueles que desejam melhores resultados na hora de realizar suas pesquisas.

3. Problema

Muitas vezes não é possível encontrar artigos específicos ou trechos dentro de artigos através da busca por palavras específicas, simplesmente por estas palavras não estarem presentes nos mesmos ou muitas vezes estando simplesmente escritas com palavras sinônimos, que contém o mesmo significado mas com outra forma de se escrever, desta maneira muitos artigos ou trechos não são encontrados e deixam de ser ferramentas de estudos.

Com visão deste problema a ideia principal torna-se aplicar uma maneira de que, durante a busca por artigos e/ou trechos ao direcionar uma palavra específica como comando de busca para localizar os mesmos, aplique-se também uma relação entre palavras sinónimos a utilizada para iniciar a busca, assim, trazendo maiores probabilidades de se obter sucesso a busca realizada.

4. Métodos

Para a realização deste experimento utilizamos métodos de Algoritmo Genético em cooperação de uma String de Busca, ao adicionar os termos necessários na String de Busca o algoritmo genético realiza uma verificação nas soluções encontradas utilizando as palavras chaves na String de Busca além de seus sinônimos e através da função fitness inicia o processo de seleção das soluções que apresentem um melhor resultado.

Algoritmo Genético é o responsável pela aplicação, para que a String de Busca percorra a busca dos melhores resultados plausíveis que são determinados pela Função Fitness, ao encontrar o melhor resultado ou atender ao requisitos pré definidos a função retornará uma String com o melhor resultado possível e seu valor total em comparação a outros resultados. O Algoritmo Genético gera mutantes através das palavras iniciais da String de Busca para então analisar as palavras e realizar a busca por artigos que contenham estas chaves em seu título e resumo, após encontrados o algoritmo utiliza de uma Função para avaliação de resultados e então poder retornar os melhores avaliados com base na String de Busca inicial.

Função Fitness: A função fitness realiza a contagem de palavras chaves no título e no resumos dos artigos encontrados, realiza a soma dos mesmo e faz a divisão pelo número de artigos carregados.

$$F() = ((\text{Quantidade de Keywords no abstract} + \text{título} + \text{palavras chaves}) / \text{número de artigos encontrados})$$

Se o valor relativo alcançar o nível desejado, considera-se como uma solução válida.

5. Resultados

Durante a execução foram criadas as Strings e seus sinônimos determinados, faltando a relação de cruzamento para a criação dos mutantes de busca e retorno dos resultados esperados, com isso não foi possível aplicar a teoria de melhoramento desejada para a aplicação, sendo assim os resultados são inconclusivos a partir da geração das strings mutantes e seus respectivos retornos.

Esperava-se como resultados o retorno de mutantes com Strings determinadas de busca que apresentariam o melhor resultado possível de artigos relacionados a uma String de Busca utilizada, assim descartaria-se artigos com pouca relação ou baixo valor Fitness ao que se é necessário.

6. Conclusão

Concluimos a partir deste experimento que a utilização de uma combinação de String de Busca e Algoritmo Genético com métodos de avaliação bem definidos pode fazer com que a busca de artigos científicos específicos para o estudo da área de conhecimento abordada pode ser amplamente melhorada, fazendo com que a relação de artigos com abordagens diferentes da requerida sejam descartados.

Como não conseguimos concluir a aplicação do algoritmo ficamos por meio desse impossibilitados de concluir o estudo e apresentar a eficácia da solução proposta.

Referências

- [1] <https://github.com/fcarlosmonteiro/hill-climbing-string-generator>
- [2] Automating Search Strings for Secondary Studies
Francisco Carlos Souza 1, Alinne Santos 1, Stevão Andrade1, Rafael Durelli 2, Vinicius Durelli 1, and Rafael Oliveira3
- [3] Base de Dados(CSV): <https://developer.ieee.org/>
- [4] <https://github.com/MarcoAntonioMartins/String-genetarion.git>