

## Uma representação hexadecimal e uma heurística VNS para seleção de *strings*

**Luiz Fernando dos Reis de Oliveira, Adria Ramos de Lyra**

Universidade Federal Rural do Rio de Janeiro (UFRRJ)  
Av. Governador Amaral Peixoto, s/n. Centro, Nova Iguaçu/ RJ. Brasil.

[lfdeoliveira@ufrj.br](mailto:lfdeoliveira@ufrj.br), [adrialyra@ufrj.br](mailto:adrialyra@ufrj.br)

Nesse trabalho, os autores focam no Problema da Sequência Mais Próxima (PSMP) e suas aplicações no contexto da Biologia Molecular. O PSMP é NP-Difícil, motivando os autores na proposição de heurísticas, com objetivo de encontrar soluções de boa qualidade.

Neste problema, utiliza-se uma métrica denominada Distância de Hamming, que representa a quantidade de posições em que duas sequências de mesmo tamanho diferem entre si. Este valor é representado por  $dh(s_1, s_2)$ .

Dado um conjunto com  $n$  sequências de entrada, todas de tamanho  $m$ , o objetivo é criar uma nova sequência, também de tamanho  $m$  que minimize a Distância de Hamming entre as sequências do conjunto.

Formalmente, temos como objetivo, a partir do conjunto de entrada  $S$ , encontrar a sequência  $s'$  com Distância de Hamming  $d$ , tal que, para todo  $s_i \in S$ ,  $1 \leq i \leq n$ ,  $dh(s_i, s') \leq d$ .

Em uma aplicação prática, o PSMP aparece no contexto do silenciamento de genes (também conhecido como RNA-interferência) onde, para criar resistência sobre uma certa população de vírus, uma possível solução é a introdução de uma sequência artificial que seja, ao menos, 95% homóloga à esta população.

Foram propostas quatro heurísticas de construção de solução inicial e um algoritmo de busca local. As implementações foram utilizadas para criar diferentes versões da meta-heurística VNS (*Variable Neighborhood Search*).

Também é proposta uma representação hexadecimal que auxilia a identificar diferentes sequências candidatas a solução inicial para o PSMP.

**Palavras Chave:** Problema da Sequência Mais Próxima, Otimização Combinatória, heurísticas.