

## Autômatos e Bioinformática

Um Grafo Acíclico Direcionado de Palavra (DAWG - do inglês: Directed Acyclic Word Graph) é um tipo de autômato finito não determinístico sem ciclos com exatamente um estado final e que reconhece uma linguagem finita. Este trabalho trata da síntese de DAWG a partir de um conjunto de dados de strings classificadas, ou seja, cada string é classificada como positiva ou negativa. Além disso, a síntese de DAWG neste trabalho será aplicada em um conjunto de dados real para classificação de proteínas amiloidogênicas. Essas proteínas são responsáveis por doenças como Alzheimer, Doença de Huntington e Diabetes Tipo 2.

- (a) Construa um programa para definir autômatos finitos não determinísticos. Sua implementação deve funcionar para qualquer tipo de autômato finito não determinístico, ou seja, deve aceitar qualquer conjunto de estados, alfabeto, função de transição, estado inicial e conjunto de estados finais. Depois do autômato finito não determinístico ser definido, deve ser possível verificar se ele aceita ou rejeita strings fornecidas. Isso deve ser feito **sem** converter o autômato finito não determinístico em um determinístico.
- (b) Desenvolva um programa para definir autômatos finitos determinísticos. Seu código deve funcionar para qualquer tipo de autômato finito determinístico, ou seja, deve aceitar qualquer conjunto de estados, alfabeto, função de transição, estado inicial e conjunto de estados finais. Além disso, sua implementação deve ser capaz de verificar se um autômato finito determinístico aceita ou rejeita uma string fornecida.
- (c) Também implemente a conversão de autômato finitos não determinísticos em autômato finitos determinísticos. Agora seu programa deve ser capaz de verificar se um autômato finito não determinístico aceita ou rejeita uma string fornecida usando a conversão para um autômato finito determinístico.
- (d) Faça a leitura do artigo **Induction of Directed Acyclic Word Graph in a Bioinformatics Task** (<http://proceedings.mlr.press/v34/wieczorek14a.pdf>) sobre síntese de DAWG. De acordo com o artigo, desenvolva um programa para construir um DAWG a partir de um conjunto de strings classificadas. Depois, utilize seu programa para gerar um DAWG a partir do conjunto de dados **waltz.txt** disponível junto deste trabalho. Em seguida, você deve testar as strings do conjunto de dados **waltzdb.csv** no DAWG obtido e verificar a porcentagem de acertos. Você deve fazer esse teste de duas maneiras: **sem** converter o DAWG em um autômato finito determinístico e o convertendo em um autômato finito determinístico. No final, em cada caso, você deve mostrar a média de tempo para processar uma string no DAWG.