

**Universidade Federal de Viçosa - Florestal**  
Ciências da Computação

**Trabalho Prático 3 – Projeto e Análise de Algoritmos**  
Trabalho 3 (similaridade de DNA)

João Victor Graciano Belfort de Andrade  
- 4694  
Guilherme Augusto Schwann Wilke  
- 4685

<b>1.Introdução</b>	<b>1</b>
<b>2.Desenvolvimento</b>	<b>2</b>
2.1. Lógica do Algoritmo	2
<b>3. Compilando o Projeto</b>	<b>2</b>
<b>4. Testes</b>	<b>3</b>
4.1 Gráfico	6
<b>5. Decisões tomadas no Projeto</b>	<b>6</b>
<b>6. Conclusão</b>	<b>6</b>
<b>7. Referências</b>	<b>6</b>

## **1.Introdução**

Esta documentação se refere ao projeto feito no Trabalho Prático 3, da disciplina ‘“PROJETO E ANÁLISE DE ALGORITMOS”’, de código CCF 330. No trabalho, foi sugerido a implementação de um algoritmo com para encontrar semelhança entre 3 DNAs, chimpanze humano e macaco, utilizando o casamento de cadeias de caracteres.

## **2.Desenvolvimento**

### **2.1. Lógica do Algoritmo**

O algoritmo proposto na descrição do trabalho foi adaptado para uma melhor execução e boa interação com o usuário, sem deixar de seguir as especificações passadas na proposta. Não foi implementado um menu de escolhas, logo o tamanho das bases a serem utilizadas no resultado do produto cartesiano e o número de produtos cartesianos selecionados são aleatórios, o resultado do produto cartesiano é então utilizado no casamento de cadeias de caracteres, então é pedido para o usuário inserir uma opção, que é qual dos casamentos descritos no trabalho ele quer utilizar, o programa cria 3 listas de inteiros(uma para cada espécie) com o mesmo valor aleatório do número de produtos cartesianos selecionados, e também é lida uma sequência aleatória do dna de cada uma das três espécies, é então feita a pesquisa das sequências de bases geradas na função de produto cartesiano, em cada sequência de dna selecionada, isso retorna 3 vetores inteiros, em cima desses valores inteiros deve ser feita a operação passada pelo professor de similaridade de cossenos, então são gerados 3 valores os de similaridade dos respectivos pares: humano-cachorro, humano-chimpanze, chimpanze-cachorro. Esses passos definem o funcionamento total do trabalho, porém para a maior assertividade foi implementada a simulação pedida pelo professor, a simulação é feita com valores que podem ser alterados pois são estáticos( constantes), a simulação basicamente faz os passos anteriores x vezes, com valores aleatórios para sequências de dna a serem lidas dos arquivos, é então feita uma média dos resultados, que se assemelha muito ao resultado real da similaridade entre os dnas.

## **3. Compilando o Projeto**

Para compilar o projeto basta digitar (\$ make all) e depois (\$ make run) dentro do diretório principal do projeto, para isso é necessário o cmake e o gcc instalados em seu computador, já para limpar basta rodar (\$make clean) e então o arquivo gerado será apagado, e poderá ser compilado novamente.

#### 4. Testes

A seguir estão os testes de cada uma das opções do menu para mostrar o funcionamento e a capacidade do algoritmo:

```
Escolha um dos 3 algoritmos para calcular o casamento de padrões de caracteres:
1 - Boyer-Moore
2 - Shift-And
3 - KMP
1
Quantas simulações deseja fazer? (max 50)
10
Fazendo simulação para 770 vezes.
HC: 0.578427,HD: 0.620212,CD: 0.604864
Fazendo simulação para 773 vezes.
HC: 1.157076,HD: 1.228455,CD: 1.195496
Fazendo simulação para 584 vezes.
HC: 1.579265,HD: 1.686402,CD: 1.639212
Fazendo simulação para 647 vezes.
HC: 2.064848,HD: 2.210201,CD: 2.139302
Fazendo simulação para 682 vezes.
HC: 2.574612,HD: 2.754745,CD: 2.667491
Fazendo simulação para 616 vezes.
HC: 3.025016,HD: 3.243934,CD: 3.142965
Fazendo simulação para 671 vezes.
HC: 3.518370,HD: 3.781664,CD: 3.657601
Fazendo simulação para 556 vezes.
HC: 3.935112,HD: 4.225504,CD: 4.093776
Fazendo simulação para 714 vezes.
HC: 4.455226,HD: 4.781466,CD: 4.632390
Fazendo simulação para 589 vezes.
HC: 4.889045,HD: 5.241156,CD: 5.084964
Fazendo simulação para 687 vezes.
HC: 5.389530,HD: 5.782452,CD: 5.600475
```

**Imagem 1-Resultados de 10 simulações executando N vezes (N aleatório) utilizando o algoritmo de Boyer-Moore.**

```
Escolha um dos 3 algoritmos para calcular o casamento de padrões de caracteres:
1 - Boyer-Moore
2 - Shift-And
3 - KMP
2
Quantas simulações deseja fazer? (max 50)
10
Fazendo simulação para 655 vezes.
HC: 0.476640,HD: 0.502288,CD: 0.500971
Fazendo simulação para 747 vezes.
HC: 1.037853,HD: 1.081500,CD: 1.078464
Fazendo simulação para 613 vezes.
HC: 1.496850,HD: 1.565325,CD: 1.560355
Fazendo simulação para 731 vezes.
HC: 2.035911,HD: 2.144883,CD: 2.119913
Fazendo simulação para 557 vezes.
HC: 2.445290,HD: 2.577265,CD: 2.542799
Fazendo simulação para 634 vezes.
HC: 2.918845,HD: 3.069000,CD: 3.029089
Fazendo simulação para 798 vezes.
HC: 3.491425,HD: 3.673575,CD: 3.626371
Fazendo simulação para 506 vezes.
HC: 3.874025,HD: 4.069382,CD: 4.025723
Fazendo simulação para 747 vezes.
HC: 4.425584,HD: 4.653826,CD: 4.599273
Fazendo simulação para 627 vezes.
HC: 4.898134,HD: 5.137932,CD: 5.081983
Fazendo simulação para 543 vezes.
HC: 5.302945,HD: 5.562731,CD: 5.496655
```

**Imagem 1-Resultados de 10 simulações executando N vezes (N aleatório) utilizando o algoritmo de Shift-And.**

```
Escolha um dos 3 algoritmos para calcular o casamento de padrões de caracteres:
1 - Boyer-Moore
2 - Shift-And
3 - KMP
3
Quantas simulações deseja fazer? (max 50)
10
Fazendo simulação para 703 vezes.
HC: 0.523731,HD: 0.560771,CD: 0.548577
Fazendo simulação para 641 vezes.
HC: 0.991271,HD: 1.063307,CD: 1.045859
Fazendo simulação para 651 vezes.
HC: 1.464872,HD: 1.570310,CD: 1.548878
Fazendo simulação para 692 vezes.
HC: 1.970008,HD: 2.103709,CD: 2.077431
Fazendo simulação para 522 vezes.
HC: 2.353180,HD: 2.524362,CD: 2.478748
Fazendo simulação para 700 vezes.
HC: 2.875315,HD: 3.069690,CD: 3.014362
Fazendo simulação para 600 vezes.
HC: 3.305978,HD: 3.534765,CD: 3.464107
Fazendo simulação para 612 vezes.
HC: 3.759345,HD: 4.004972,CD: 3.938911
Fazendo simulação para 739 vezes.
HC: 4.304018,HD: 4.581600,CD: 4.507948
Fazendo simulação para 559 vezes.
HC: 4.731292,HD: 5.030567,CD: 4.949049
Fazendo simulação para 643 vezes.
HC: 5.208400,HD: 5.535619,CD: 5.447600
```

**Imagem 1-Resultados de 10 simulações executando N vezes (N aleatório) utilizando o algoritmo KMP.**

## **5. Decisões tomadas no Projeto**

A principal decisão feita no projeto foi a de como realizar o algoritmo de casamento de cadeias de caracteres, optamos por pesquisar mais sobre os algoritmos e buscar em diversas fontes como implementá-los.

## **6. Conclusão**

De forma geral as diretrizes passadas pelo professor de CCF-330 foram seguidas para a realização do trabalho, tanto para projeção e execução do código quanto para confecção deste documento. Os resultados foram os esperados de acordo com as diretrizes, tentando conduzir uma abordagem mais prática possível com o usuário. Deste modo o trabalho foi conduzido com criatividade e seriedade para poder-se ter uma boa ideia de funcionamento, aplicação e execução de um algoritmo com casamento de cadeias de caracteres, podendo implementar os 3 algoritmos apresentados na matéria, dessa forma, cumprindo todas as tarefas extras apresentadas pelo docente.

## **7. Referências**

Instruções para realização do trabalho [1].

Instruções para realização da documentação [2].  
casamento de cadeias–trabalhosfeitos [3]  
Processamento de Cadeias de Caracteres-UFOP.[4]  
Cosine Similarity-wikipedia[5]

[1] Diretrizes básicas para as documentações da disciplina. Fornecida pelo professor Daniel através do Moodle.

[2] Trabalho 2. Fornecida pelo professor Daniel através do Moodle.

[3]<http://www.decom.ufop.br/menotti/paa111/slides/aula-ProcCadCarac.pdf>

[4]<https://www.trabalhosfeitos.com/ensaios/Casamento-De-Cadeias/45721428.html>

[5][https://en.wikipedia.org/wiki/Cosine\\_similarity](https://en.wikipedia.org/wiki/Cosine_similarity)