

ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

PROGRAMAÇÃO DE COMPUTADORES II

Prática 30: Trabalho de Arquivo Texto

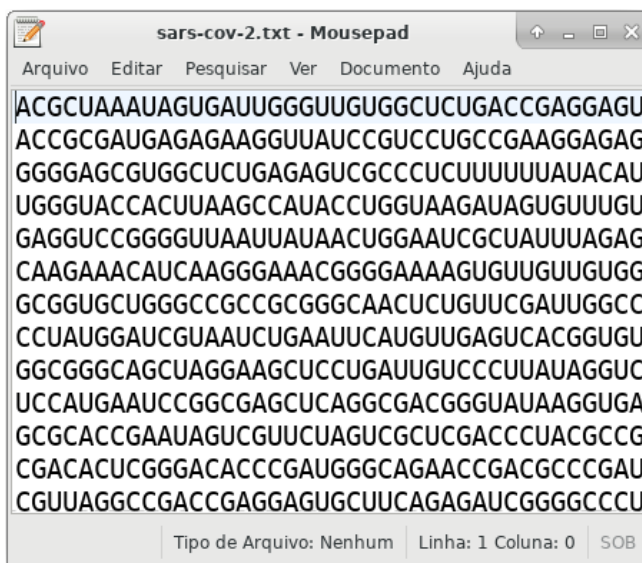
A atividade proposta utiliza-se de informações totalmente hipotéticas, e que não possuem valor científico dentro da biologia. O único intuito do trabalho é colocar o aluno perante em uma situação real onde necessite utilizar os conceitos abordados durante as aulas, sobre arquivos texto.

1. Introdução

Um laboratório brasileiro está efetuando um estudo sobre o código genético do vírus SARS-CoV-2, causador da Covid-19. Dentre as diversas informações importantes desse estudo, uma delas é a identificação de determinados padrões dentro da sequência genética mapeada para SARS-CoV-2.

2. Atividade Desenvolvida

Com intuito de acelerar o processo do estudo desse novo vírus, os pesquisadores do laboratório resolveram contratar um programador capaz de varrer o arquivo texto que contém o mapeamento do código genético do SARS-CoV-2. Como se trata de um vírus, seu genoma é constituído por uma fita simples de RNA, ou seja, sua sequência genética só envolve as bases nitrogenadas A (Adenina), U (Uracila), C (Citosina) e G (Guanina). O código genético mapeado é composto por uma única sequência contendo milhares de bases nitrogenadas. Contudo, para que fosse possível organizar essa sequência (quantidade de caracteres) dentro de um arquivo texto, ela foi quebrada em várias linhas de 40 bases nitrogenadas (caracteres). Uma imagem, da parte inicial do arquivo, é apresentada a seguir ([baixar Linux](#) / [baixar Windows](#)):



```
sars-cov-2.txt - Mousepad
Arquivo  Editar  Pesquisar  Ver  Documento  Ajuda

ACGCUAAAUAGUGAUUUGGUUGUGGCUCUGACCGAGGAGU
ACCGCGAUGAGAGAAGGUUAUCCGUCCUGCCGAAGGAGAG
GGGGAGCGUGGCUCUGAGAGUCGCCUCUUUUUUUAJACAU
UGGGUACCACUUAAGCCAUACCUUGGUAAGAUAGUGUUUGU
GAGGUCCGGGUUAAUUAUAACUGGAAUCGCUAUUUAGAG
CAAGAAACAUAAGGGAAACGGGAAAAGUGUUGUUGUGG
GCGGUGCUGGGCCGCGCGGGCAACUCUGUUCGAUUGGCC
CCUAUGGAUCGUAUUCUGAAUUAUGUUGAGUCACGGUGU
GGCGGGCAGCUAGGAAGCUCCUGAUUGUCCUUAUAGGUC
UCCAUGAAUCCGGCGAGCUCAGGCGACGGGUAUAAGGUGA
GCGCACCGAAUAGUCGUUCUAGUCGCUACCCUACGCCG
CGACACUCGGGACACCCGAUGGGCAGAACCGACGCCCGAU
CGUUAGGCCGACCGAGGAGUGCUUCAGAGAUCGGGGCCCU
```

Tipo de Arquivo: Nenhum | Linha: 1 Coluna: 0 | SOB



A partir do arquivo os pesquisadores podem buscar padrões que possibilitam identificar características específicas sobre a natureza do SARS-CoV-2. Para tal, a aplicação criada deve ser capaz de, ao receber uma sequência de base nitrogenada, identificar a quantidade de vezes que ela aparece no mapa genético do vírus. O algoritmo deve buscar a sequência recebida ignorando as quebras linha do arquivo. Em outras palavras, a sequência recebida pode começar numa linha e terminar na seguinte. Ao final, a aplicação deve apresentar a quantidade de vezes em que a sequência buscada foi encontrada, como apresentado a seguir:

```
Terminal - ./a.out
Arquivo  Editar  Ver   Terminal  Abas  Ajuda
-----
- Bases Nitrogenadas => A U C G -
-----
- Digite a sequência: UUCG
- Buscando(UUCG)...
- Total de ocorrências [UUCG] = 95

Deseja efetuar uma nova análise (s) ou (n)? s

Terminal - ./a.out
Arquivo  Editar  Ver   Terminal  Abas  Ajuda
-----
- Bases Nitrogenadas => A U C G -
-----
- Digite a sequência: AAGUC
- Buscando(AAGUC)...
- Total de ocorrências [AAGUC] = 15

Deseja efetuar uma nova análise (s) ou (n)? n
```