

UNIVERSIDADE FEDERAL DO PIAUÍ - CAMPUS SHNB
SISTEMAS DE INFORMAÇÃO
ALGORITMOS E PROGRAMAÇÃO II
Trabalho Prático

1 Sumarizador de Notícias

Atualmente as pessoas estão muito ocupadas e têm pouco tempo de acompanhar notícias dos principais portais de comunicação. Nesse contexto, uma aplicação útil são sumarizadores de texto, que mostram de uma forma bem sucinta e criativa o conteúdo de um texto, ajudando o usuário rapidamente verificar se o conteúdo lhe interessa.

O objetivo desse trabalho prático é desenvolver um programa sumarizador de conteúdo básico, que funciona via console do computador. O programa receberá como entrada um conteúdo de texto e exibirá na tela um sumário do conteúdo. O sumário consiste na lista de palavras chaves do texto, ou seja, as n palavras mais frequentes, iniciando da palavra mais frequente para a menos frequente. Por sua vez, o conteúdo de texto é um arquivo de texto (puro).¹ Opcionalmente o conteúdo de texto também pode ser o endereço para uma página web, ou seja, um arquivo HTML. Essa opção valerá pontos extras.

As seções a seguir descrevem os detalhes de como o programa deve ser desenvolvido e apresentado.

2 Instruções sobre Implementação

O programa deve ser implementado na linguagem C. Para padronizar a correção do trabalho prático, o código deve ser compilado e executado em terminal do sistema Linux ou Windows nas máquinas dos laboratórios 813 ou 814. A execução do programa será via linha de comando da seguinte forma:

Sistemas Linux

```
./tp <parametro1> <parametro2> <parametro3> <parametro4> <parametro5>
```

Sistemas Windows

```
tp <parametro1> <parametro2> <parametro3> <parametro4> <parametro5>
```

Os parâmetros acima devem ter respectivamente as seguintes informações:

1. o tipo de conteúdo (1-website, 2-arquivo de texto puro);
2. o caminho para o website ou arquivo de texto;
3. o tamanho do resumo (a quantidade de palavras);
4. o caminho para o arquivo de texto com *stopwords*;
5. o caminho para o arquivo de texto de saída (resumo);

O arquivo de *stopwords* contém palavras “irrelevantes” que devem ser desconsideradas pelo programa. Essas palavras devem estar separadas por uma vírgula e um espaço. Para fim de testes iremos considerar as seguintes *stopwords*:

¹Não pode ser arquivos de editores de texto como MS Word ou Open Office.

a, as, o, os, um, uns, uma, umas, para, de, da, do, das, dos, se, então, ele, ela, eles, elas, à, ou, por, que, com, também, the, a, an, at, by, from, how, of, on, that, to, for, in, out, I, he, she, it, they, them, him, her, do, will, or, with, also

Para o arquivo de saída, o programa deverá exibir os resultados no seguinte formato:

```
palavra1 valor1
palavra2 valor2
...
palavran valorn
```

Nessa saída, palavra_{*i*} e valor_{*i*} representam respectivamente uma palavra do texto e a frequência dessa palavra em ordem decrescente, onde a posição *n* contém a palavra menos frequente, ou seja, dentre as *n* palavras a que ocorre em menor quantidade.

3 Documentação

A documentação deve conter pelo menos quatro seções, organizadas da seguinte forma:

- Uma visão geral da solução proposta
- A organização da solução, ou seja, como a solução foi modularizada.
 - Não inclua código em C!
 - Use algoritmos genéricos (exemplo, portugol) ou diagramas de fluxos para descrever funções e modularização utilizada.
- Análise de custo da solução proposta.
- Conclusão curta em um parágrafo.

A análise de custo deve explicar quantas vezes a função mais importante do programa (ou trecho dessa função) é executada em relação ao tamanho da sentença. Gráficos e/ou tabelas podem ser utilizados nessa análise.

A documentação deve conter no máximo 7 páginas, seguindo os modelos doc ou latex a serem disponibilizados no compartilhamento. A partir desses modelos deve ser gerado um arquivo PDF para a entrega.

4 Avaliação

O trabalho prático pode ser feito em grupos de até **3 componentes** e os pontos desse trabalho terão a seguinte organização:

- Funcionamento correto do programa (60% dos pontos)
- Documentação (40% dos pontos)

A documentação (um arquivo PDF) e códigos fonte do trabalho devem ser compactados em arquivo do tipo zip e submetidos no sistema Sigaa. Após a submissão, será agendado entrevistas com os grupos. O objetivo da entrevista é verificar se cada membro do grupo participou do desenvolvimento do trabalho. **Caso seja observado falta de conhecimento do trabalho entregue, haverá reduções da nota do membro, podendo alcançar nota 0 o membro que desconhecer totalmente o trabalho.**

Atenção: cópias de código ou soluções idênticas do problema, levará a **nota zero** para todos os membros dos grupos envolvidos nas cópias e soluções.

5 Datas

- Submissão no Sigaa na sexta-feira (17/03/2023) até as 23:59.
- Entrevistas a serem agendadas posteriormente com grupos.

Boa sorte!