

## Trabalho Prático

### Problema: Indexação de palavras em textos

O problema proposto envolve localizar as linhas que as palavras foram digitadas em um texto qualquer. Por exemplo, considerando o texto a seguir, criado no Mussum Ipsum (<https://mussumipsum.com/>), a palavra *amistosis* está na linha 2.

*Mussum Ipsum, cacilds vidis litro abertis. Nullam volutpat risus nec leo commodo, ut interdum diam laoreet. Sed non consequat odio. In elementis mé pra quem é amistosis quis leo. Suco de cevadiss, é um leite divinis, qui tem lupuliz, matis, aguis e fermentis. Detrahit consequat et quo num tendi nada.*

*Praesent vel viverra nisi. Mauris aliquet nunc non turpis scelerisque, eget. Mé faiz elementum girarzis, nisi eros vermeio. Quem manda na minha terra sou euzis! Posuere libero varius. Nullam a nisl ut ante blandit hendrerit. Aenean sit amet nisi.*

O seu programa deve então:

- criar um mecanismo eficiente para ler o texto e salvar em qual linha cada palavra foi encontrada
- permitir que o usuário busque palavras do texto, retornando para o usuário a linha (ou linhas) onde ela foi encontrada.

Como estrutura para salvar as palavras, você deve utilizar uma tabela hash com endereçamento fechado (cada posição da tabela é uma lista). A chave para salvar na tabela deve ser a própria palavra, e caso duas palavras iguais sejam encontradas, ambas as entradas devem ser inseridas na tabela, para registrar as linhas onde foram encontradas.

A função de transformação deve ser o resto da divisão entre a soma dos valores dos caracteres da palavra em ASCII e o número de entradas da tabela, como a equação abaixo.

$$h = k \% t, \quad k = \sum_{i=0}^{i=\text{strlen}(p)} p[i] \quad (1)$$

Sendo  $p$  a *string* a ser inserida na tabela,  $\text{strlen}(p)$  o tamanho desta *string*, e  $t$  o tamanho da tabela hash.

### Itens a serem avaliados no trabalho

Tabela hash com inserção funcionando e leitura do arquivo de texto	60% da nota
Busca das palavras funcionando	+10%
Listagem correta da linha quando a palavra for encontrada só uma vez no texto	+10%
Listagem correta das linhas quando a palavra for repetida no texto	+10%
Impressão da tabela hash com todas as palavras, seu valor de hash calculado e a linha encontrada.	+10%

## Observações sobre a entrega:

A não observação destes itens zera a nota do trabalho!

- (a) O trabalho pode ser feito em grupo de no máximo três pessoas.
- (b) A parte de implementação do trabalho deverá ser realizada utilizando a linguagem C e ser entregue em um único arquivo compactado, contendo o código-fonte e um arquivo README detalhando como compilar e executar o programa.
- (c) A entrada do programa deve ser via terminal, recebendo como parâmetro somente os nomes do arquivo com o texto de entrada, e a seguir solicitar as palavras a serem buscadas, ou que se digite 0 para sair.
- (d) Para cada etapa, você deve implementar o seu próprio TAD Dicionário, utilizando a tabela hash.
- (e) Nesse zip não deve haver arquivos executáveis.
- (f) Incluir pdf da parte escrita no zip.
- (g) **Data de entrega: 10/12/2019.** Não há possibilidade de alteração no prazo de entrega devido ao prazo para correção e fechamento de notas do semestre.
- (h) Valor: 25 pontos

### O que deve ser entregue:

- Documentação do trabalho. Em entre outras coisas, a documentação deve conter:
  1. Introdução: descrição do problema a ser resolvido e visão geral sobre o funcionamento do programa.
  2. Implementação: descrição sobre a implementação do programa e principalmente a análise de complexidade dos algoritmos implementados.
  3. Conclusão: comentários gerais sobre o trabalho e as principais dificuldades encontradas em sua implementação.
  4. Bibliografia: bibliografia utilizada para o desenvolvimento do trabalho, incluindo sites da Internet, se for o caso.
- Além disso, neste trabalho deve ser enviado ao professor os códigos-fonte do trabalho. A entrega deverá ser feita via Moodle, seguindo as diretrizes informadas no início da descrição deste trabalho.

### Comentários Gerais:

- Comece a fazer este trabalho logo, enquanto o problema está fresco na memória e o prazo para terminá-lo está tão longe quanto jamais poderá estar.
- Procurem o monitor da disciplina! Ele pode ajudar naquele momento em que o TP esteja mais complicado.
- Clareza, indentação e comentários no programa também vão valer pontos.
- Trabalhos copiados serão penalizados com a nota zero.