

Universidade do Estado de Santa Catarina (UDESC)
Curso: Tecnologia em Análise e Desenvolvimento de Sistemas
Disciplina: Linguagem de Programação (LPG0001)
Prof. Rui J. Tramontin Jr.

Trabalho: Implementação de um Indexador

O objetivo deste trabalho é implementar um programa que faça a **indexação** de um arquivo de texto, tendo como saída um arquivo texto contendo uma lista de palavras e as linhas onde ocorrem no texto original. A figura a seguir apresenta um exemplo com os arquivos de entrada e de saída.

Arquivo de entrada

```
1 To be or not to
2 be, that is
3 the question.
```

Arquivo de saída

```
be
- Ocorrências: 2
- Linhas: 1, 2

is
- Ocorrências: 1
- Linhas: 2

not
- Ocorrências: 1
- Linhas: 1

or
- Ocorrências: 1
- Linhas: 1

question
- Ocorrências: 1
- Linhas: 3

that
- Ocorrências: 1
- Linhas: 2

the
- Ocorrências: 1
- Linhas: 3

to
- Ocorrências: 2
- Linhas: 1

Total de palavras indexadas: 8
Palavras que mais ocorrem: be, to (2 vezes)
```

Requisitos:

- 1) O índice deve ser implementado como um vetor, onde cada elemento é uma *struct* que contém os seguintes campos:
 - palavra (*char**);
 - número de ocorrências (*int*);
 - vetor de ocorrências (*int*), contendo as linhas nas quais palavra ocorre;
 - quantidade de valores no vetor de ocorrências (nem sempre é igual ao número de ocorrências → veja o caso do “to” no exemplo).
- 2) A *palavra* e o *vetor de ocorrências* devem ser alocados dinamicamente; o vetor de ocorrências deve ser realocado ao longo do processamento;
- 3) As entradas da lista devem ser inseridas em ordem alfabética conforme o campo “*palavra*”; o mesmo vale para a lista de ocorrências; alternativamente, as palavras podem estar fora de ordem, mas as entradas do índice devem ser ordenadas ao final;
- 4) O processamento do arquivo de entrada tem os seguintes requisitos:
 - São consideradas palavras somente uma sequência de letras (maiúsculas ou minúsculas), ou seja, todos os demais caracteres deverão ser ignorados;
 - Cada palavra deve ter suas letras convertidas para o minúsculo;
 - O índice não pode conter palavras repetidas;
 - Caso a palavra já exista no índice, o número de ocorrências é incrementado, e a linha na qual ela ocorre é incluída (se necessário);
 - A lista de ocorrências não deve conter linhas repetidas, mesmo se a palavra ocorrer mais de uma vez na mesma linha (veja a palavra “to” no exemplo);
 - Cada palavra deve ser uma *string* alocada dinamicamente.

Critérios de avaliação:

- **Coletivos:** execução correta e alinhamento com o que foi solicitado neste enunciado.
- **Individuais:** domínio do assunto e do programa implementado.

Informações importantes:

- **Equipe:** 2 a 3 alunos.
- **Entrega:** por e-mail, até o dia **28/11**.
- **Apresentação:** até o dia **28/11**, em horário a ser previamente agendado pelas equipes.