

ED23S - Estrutura de dados

4ª Prática - Tabela de dispersão (hash table)

Em grupos de 2 ou 3 alunos faça a implementação de um programa que compute estatísticas de um arquivo de texto, B .

Seu programa receberá dois arquivos:

- um arquivo, A , contendo palavras chaves, uma por linha.
- um arquivo, B , contendo um texto em várias linhas, cada linha com várias palavras.

Seu programa procurará palavras no texto de entrada e computará as seguintes estatísticas:

1. **linhas:** O número total de linhas do arquivo de entrada.
2. **palavras:** O número total de palavras no arquivo.
3. **quantidade de cada palavra chave:** Retorne o número de ocorrências de cada palavra chave.
4. **palavra chaves:** O número total de palavras chaves.

Para os itens 3 e 4 você deve implementar e utilizar uma **tabela hash**.

Um programa funcional e bem organizado receberá nota 7.0, as seguintes funcionalidades complementarão a nota:

1. (1,5) Ignorar pontuação no texto (e.g. ',', '.', '!', '?'), note que '#' não é um carácter de pontuação e portanto 'ola' e 'ola#' são duas palavras diferentes. E 'quebra-cabeça' é uma única palavra.
2. (1,5) Verificar que o arquivo de palavras chaves, A , está formatado corretamente, veja que você deve ignorar linhas em brancos, mas deve imprimir erros caso:
 - (a) Haja duas ou mais palavras numa linha
 - (b) Existência de caracteres de pontuação
 - (c) Qualquer outro erro que julgar necessário

Você deve submeter seu código fonte num arquivo zip no Moodle. No arquivo `main.c` incluir um comentário no início do arquivo com:

- Os nomes, RAs e número de chamada dos integrantes da equipe;
- Indicar quais funcionalidades complementarem foram implementadas.

Para cada funcionalidade complementar inclua arquivos de exemplos (texto (B) e palavra chave (A)).