

Hash de Palavras

Professora Patrícia Dockhorn Costa, patricia.d.costa@ufes.br

Descrição

Nesta prova você deve implementar um programa para contar a frequência das palavras em um arquivo texto, utilizando uma **Tabela Hash**.

Em um primeiro momento, o programa deve:

- Abrir o arquivo de entrada (“entrada.txt”) e ler palavras;
- Armazenar cada palavra (diferente) na Hash (considere a própria palavra como chave de busca);
- Incrementar a contagem de palavras já inseridas na hash;
- Contabilizar o número de palavras distintas e o número total de palavras (use a hash para isso);
- Descobrir a palavra com maior frequência (use a hash para isso).

Na segunda etapa, deve-se:

- Alocar um vetor que tenha como tamanho, o número de palavras distintas;
- Ordenar o vetor, por ordem crescente de frequência das palavras (use a *quicksort* da biblioteca padrão);
- Abrir o arquivo de saída (“saida.txt”) e escrever as palavras ordenadas por frequência.

Considere o arquivo a seguir, como exemplo:

```
exemplo de arquivo de entrada de dados para a prova de estruturas de dados.
```

O arquivo de saída deve conter uma lista de palavras ordenadas, suas frequências, o número de palavras distintas, o número total de palavras e a palavra de maior frequência entre as palavras. Para o exemplo acima, a seguinte saída é esperada:

```
a 1
arquivo 1
entrada 1
estruturas 1
exemplo 1
para 1
prova 1
dados 2
de 5
-----
9 palavras distintas encontradas
14 palavras no total
de é a palavra de maior frequência (5)
```

Regras importantes:

Monte a hash e só depois a utilize para achar o número de palavras distintas, o número total de palavras e a palavra de maior frequência entre as palavras. Ou seja, NÃO calcule o número de palavras e frequências durante a leitura do arquivo;

1. Utilize tipos de dados estruturados;
2. Organize sua prova em tad's;
3. Não utilize variáveis globais;
4. O arquivo com a função main deve ter o seu nome: <nome>.c;
5. Libere toda a memória alocada (use valgrind).

BOA PROVA!