EP4 - Freqüência de palavras

Entrega: 13/11/2006

Sua tarefa é escrever um programa que dado um arquivo texto e um inteiro k, mostra as k palavras mais comuns no arquivo e número de ocorrências de cada palavra em ordem decrescente de freqüência.

Por simplicidade, uma palavra é uma seqüência contígua de caracteres entre 'A-Z' e 'a-z' (inclusive) que não contém nenhum outro caractere (note que a definição de palavra aqui é diferente da definição no wc do ep1). Qualquer outro caractere DEVE ser interpretado como um separador de palavras. Ao contar o número de palavras você DEVE assumir que letras maiúsculas e minúsculas são iguais, ou seja, 'Casa' e "caSa" devem ser contadas como duas ocorrências de uma mesma palavra. Você pode assumir que toda palavra possui no máximo 40 caracteres. (Em algum momento você vai ter que usar um algoritmo de ordenação eficiente).

Você pode achar mais fácil fazer um programa que supõe que o texto tem no máximo 10.000 caracteres. O ideal é que o seu programa não impusesse tal limite artificial. Você pode achar mais fácil fazer um programa que considera "Casa" e "caSa" como palavras diferentes. Você pode corrigir estes pontos mais a frente, quando seu programa estiver "quase pronto".

Exemplo de entrada e saída:

```
ep4.exe bible11-PartOfChapter1.txt 10
   2500 and
   1735 the
    908 of
    437 to
    415 he
    363 his
    362 in
    358 that
    357 unto
    352 i
```

O arquivo usado no teste acima pode ser baixado de

`http://www.ime.usp.br/~fabricio/bible11-PartOfChapter1.txt`.

Uma boa fonte de arquivos para testar seu programa é o site `http://www.gutenberg.org/`.

No decorrer do tempo, dicas serão lançadas no panda. Portanto fique atento!