TSA-1:Trabalho em Sala de Aula 1

Tema: Módulo Fundamentos

A. Enunciado:

Bag of Words é um modelo de representação simplificado utilizado em sistemas de recuperação de informações. Trata-se do levantamento estatístico de termos que ocorrem em um determinado texto (https://en.wikipedia.org/wiki/Bag-of-words_model).

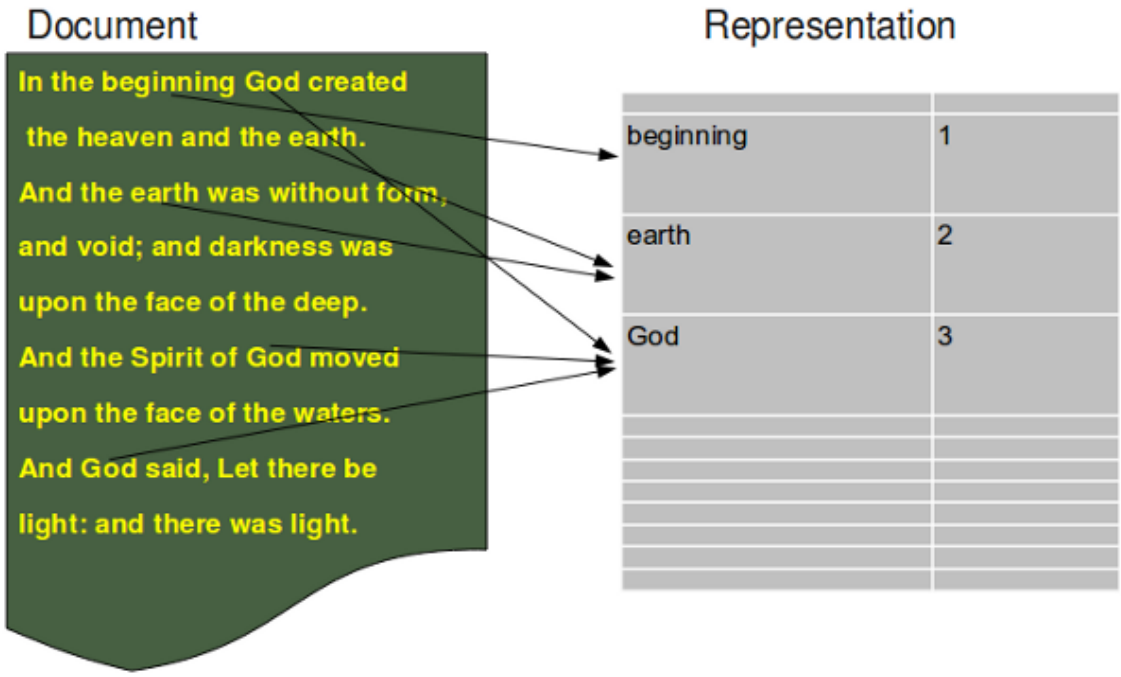

Figura 1. Bag of Words

Faça um programa que leia um arquivo contendo um texto e grave um arquivo de saída (com o nome ***bagofwords.csv***) contendo: ***palavra; quantidade; percentual***, onde ***palavra*** é uma das palavras encontradas no arquivo de entrada, ***quantidade*** é um número inteiro representando a quantidade de ocorrências desta palavra ao longo do texto e ***percentual*** é um valor real (com 2 casas decimais) contendo o percentual de ocorrência desta palavra em relação ao tamanho total do texto (quantidade de palavras).

Quadro 1. Arquivo ***bagofwords.csv*** (***palavra; quantidade; percentual***) com 200 palavras

casa; 2; 1.00%
rua; 10; 5.00%
.....
.....
estrada; 50; 25.00%

B. Requisitos:

a. O arquivo ***bagofwords.csv*** deve estar ordenado pelo campo **quantidade**.
b. Pontuações e Palavras com menos de 3 letras devem ser descartadas;
c. O sistema deve funcionar para qualquer arquivo contendo um texto;
d. **O trabalho deve ser entregue na plataforma Moodle dentro do prazo estipulado no mesmo. Não serão aceitos trabalhos entregues por outra via e/ou fora do prazo;**

C. Critérios de Avaliação:

a. Arquivo ***bagofwords.csv*** com dados de todas as palavras do arquivo de entrada: 5,0;
b. Arquivo ***bagofwords.csv*** ordenado pelo campo **quantidade**: 2,0;
c. Solução usando vetores de struct: 2,0;
d. Solução usando subrotinas ou Orientado a Objetos: 1,0;