Universidade Federal de Uberlândia - UFU  
GSI024 – Organização e Recuperação da Informação – 2017/01

Ítalo Louis Araujo, 11221BSI220  
Rogger Fernandes Barsanulfo, 11221BSI244

Etapa 1 – Projeto Final (Escrita/leitura de arquivos e contagem de termos em um documento)

# Descrição:

O projeto consiste em ler um arquivo, realizar a contagem de termos desse arquivo e realizar a escrita de termos e quantidade em outro arquivo.

# Linguagem:

O programa foi construído na linguagem Java.

# O algoritmo:

O algoritmo foi construído com os seguintes objetos:

* Termo
  + Classe responsável por declarar os atributos utilizados como nome e quantidade para o objeto termo.
  + Contém métodos de acesso aos atributos.
  + Contém os métodos hashcode e equals para realizar a comparação.
* FileParser
  + Contém o método toList
    - Responsável em realizar a leitura do arquivo usando a classe Scanner.
    - Realiza a iteração por cada palavra do arquivo retirando pontuação e padronizando as palavras como minúsculas.
    - Condiciona que a palavra tem que ter mais de 3 letras.
    - Realiza a soma da quantidade de termos que possui uma palavra.
    - Adiciona em um array de termos as palavras e quantidades lidas daquele arquivo lido.
    - Fecha o arquivo.
  + Contém o método writeFile
    - Responsável por escrever em outro arquivo a lista criada pelo array de termos.
* Main
  + Método principal que realiza chamada dos métodos do FileParser.

Após realizar o processamento do arquivo tivemos o seguinte arquivo como saída:



# Gráfico:

O gráfico a seguir consiste nos dados do arquivo gerado onde o eixo X é a ordem da palavra na lista classificada por frequência e o eixo Y é a frequência da palavra.

Como podemos verificar o gráfico possui uma curva decrescente, em que no início temos a palavra brasil com maior número de frequência, essa frequência vai caindo sendo que no intervalo entre nosso e liberdade a frequência se torna constante assim como nos intervalos entre esperança e solo.