

Trabalho Prático – Organização e Recuperação da Informação

Prof. André R. Backes

Instruções

- **Número de integrantes por grupo 3 pessoas. Grupos com mais ou menos pessoas somente serão aceitos se não houver número suficiente de alunos para formar outro grupo.**
- **Atenção: alguns grupos poderão ser sorteados para apresentar o trabalho quando houver dúvidas quanto a sua implementação.**

Data de Entrega:

10/09/2024

Forma de Entrega:

O trabalho deverá ser entregue por e-mail: **arbackes@yahoo.com.br**. Deverá ser enviado apenas o código fonte do programa. No campo **assunto** do e-mail, colocar **TrabalhoORI**.

No arquivo **main** do programa, e no corpo do e-mail, deverá constar os nomes dos integrantes do grupo.

Problema

O grupo deverá escrever um programa em linguagem C que leia um arquivo contendo postagens extraídas do Twitter:

- Cada linha do arquivo corresponde a uma postagem individual.
- Cada linha contém 3 informações separadas por vírgulas. A última é a postagem
- **Exemplo: 14,0, jb isnt showing in australia any more!**

Em seguida, o grupo deverá implementar um buscador de postagens baseados em palavras. Basicamente, o programa deverá pedir ao usuário para digitar uma lista de palavras e exibir todas as postagens que contém as palavras digitadas pelo usuário. O usuário poderá usar os operadores AND e OR para NOT para compor a sua busca:

(usa **AND NOT(winner)**) OR (brazil **AND** carnaval)

Ao final da consulta, o usuário poderá fazer uma nova consulta ou finalizar o programa.

Para a implementação, utilize as implementações das estruturas vistas em aula: **github.com/arbackes/**

Dicas:

- Para a indexação das postagens, utilize uma Tabela Hash para armazenar cada palavra das postagens.
- Para cada palavra, armazene o RRN da postagem no arquivo de modo a obter rápida indexação.
- Você pode usar uma estrutura do tipo Set para fazer a união e intersecção dos resultados.
- Desconsidere pontuações e espaços no início e fim ao armazenar as postagens.

Instruções:

O trabalho será avaliado principalmente levando em consideração:

- 1) Realização das tarefas do trabalho.
- 2) Representação correta da entrada e saída dos dados.
- 3) Uso correto das variáveis e estruturas de dados.
- 4) Uso adequado dos conceitos aprendidos em sala (modularização, hash, árvores, etc, quando for o caso).
- 5) Boa indentação e uso de comentários no código. Evite utilizar comentários excessivamente.