



Disciplina: Estrutura de Dados I
Professor: Ivairton M. Santos

Trabalho 9 – Análise do DataSet de Músicas

Considere o *dataset* (Music Dataset : 1950 to 2019 - Lyrics and Metadata from 1950 to 2019*) disponibilizado para realização deste trabalho.

A base de dados consiste de um arquivo CSV com 31 dados (colunas) sobre letras de músicas entre os anos de 1950 e 2019, com pouco mais de 28 mil registros de letras musicais.

O trabalho tem como objetivo comparar estruturas de dados diferentes, deverá ser implementado em linguagem C e será dividido em 3 partes:

Parte 1 – Carregamento da base de dados:

Inicialmente o arquivo do *dataset* deverá ser carregado totalmente em memória, numa estrutura de dados a ser escolhida/definida pela dupla. O processo de abertura e carregamento dos dados deverá ser medido em relação ao tempo.

Parte 2 – Processamento dos dados:

Com os dados carregados em memória deverá ser realizada uma análise em todas as tuplas da base de dados de modo que: um dos campos da base de dados chama-se “danceability”, ele consiste de um índice entre 0 a 1 (número real).

Todos os registros com *danceability* igual ou superior a 0,5 devem ser armazenados em uma nova estrutura de dados. Após registrar estes dados, deverá ser contabilizado o total de letras musicais que têm “genre” (gênero) igual a “pop”.

Este processo irá ocorrer por duas vezes, na primeira vez os dados selecionados por terem “danceability” igual ou maior que 0,5 deverão ser armazenados numa lista, na segunda vez os dados serão armazenados numa estrutura escolhida por vocês.

Parte 3 - Análise:

Deverá ser medido o tempo necessário para realizar a Parte 1, a Parte 2 e o total da aplicação. Também deverá ser medido o total de memória utilizado para a execução do contexto com lista e posteriormente com a estrutura de dados escolhida.

Escreva um relatório, justifique a escolha da estrutura de dados utilizada na Parte 1 e na Parte 2 e

* <https://www.kaggle.com/datasets/saurabhshahane/music-dataset-1950-to-2019>

apresente os dados medidos, atendendo aos seguintes requisitos:

- Tempo:
 - Tempo da Parte 1 = ____ segundos/minutos
 - Tempo da Parte 2 = ____ segundos/minutos
 - Tempo total = ____ segundos/minutos
- Memória:
 - Memória total para o contexto 1 (lista) = ____ MBytes
 - Memória total para o contexto 2 (est. dados escolhida) = ____ MBytes
- Contabilidade:
 - Total de letras musicais com “danceability” igual ou maior que 0,5 = ____
 - Total de letras musicais com “danceability” ($\geq 0,5$) e gênero pop = ____

Observações a respeito do trabalho:

- Escolher as estruturas de dados é importante. Uma má escolha representa um trabalho de qualidade ruim.
- Observe os campos disponíveis no dataset, em especial a primeira coluna, que possui um código associado a cada letra.
- Será avaliado a qualidade do código (legibilidade), modularização, implementação das estruturas de dados, tempo de execução, consumo de memória, facilidade de uso, relatório e apresentação do trabalho.