

TRABALHO PRÁTICO PARTE 1

PRAZO PARA ENTREGA: 15/06/2021

1. Introdução

Spotify é um serviço de streaming de música e podcast mais popular do mundo. O serviço se baseia no pagamento de royalties considerando o número de audições de um artista como proporção total de músicas ouvidas no serviço. A base de dados disponível em <https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks> reúne informações de aproximadamente 600.000 músicas lançadas entre 1922 até abril de 2021. Além disso, agrupa informações de aproximadamente 1.1 milhões de artistas. Esses dados podem ser encontrados no link supracitado, nos arquivos *tracks.csv* e *artists.csv*. No mesmo link, encontra-se os atributos de cada uma dessas entidades.

2. Objetivos

O objetivo principal deste trabalho é dar ao aluno a oportunidade de manipular arquivos através de uma base real de dados. Espera-se que o aluno seja capaz de:

- Manipular adequadamente arquivos em modo texto.
- Definir estruturas de dados para representar as entidades
- Gravar e manipular registros em arquivos binários.

3. Desenvolvimento

As etapas para o desenvolvimento do trabalho são descritas abaixo.

Etapas 1: Processamento dos dados

Sua primeira tarefa será fazer a leitura dos dados dos arquivos *tracks.csv* e *artists.csv* de forma correta, armazenando-os em estruturas de dados apropriadas. Você é livre para escolher a forma de representação dos seus dados, mas lembre-se de que você será avaliado em função da qualidade das suas escolhas. Na sequência, você deverá salvar esses conjuntos de registros em arquivos binários intitulados *tracks.bin* e *artists.bin*.

Etapas 2: Importação de Registros e Módulo de Teste

Desenvolver uma função que deverá importar conjuntos de N registros aleatórios do arquivo binário, armazenando-os em um arranjo de registros. Deve ser implementado também um módulo de testes simples, que realiza algumas operações para garantir que as funções básicas do programa entregam resultados corretos. Esse módulo deve conter funções para exibir as saídas da operação de importação dos N registros aleatórios;

Para estes testes, você deverá oferecer ao usuário a possibilidade de exibir a saída no console ou salvá-la em um arquivo. Para a saída em console, utilize $N = 10$. Para a saída em arquivo texto, utilize $N = 100$.

O módulo deve ser aplicado para as duas bases de dados.

4. Relatório

Você deverá confeccionar um relatório sobre o trabalho desenvolvido. Este relatório deve conter, obrigatoriamente, os seguintes itens:

- Detalhamento das atividades realizadas por cada membro do grupo;
- Explicações das estruturas de dados implementadas, bem como as razões para as decisões de implementação;
- Toda e qualquer referência utilizada no desenvolvimento do trabalho.

Note que o relatório deve ser formal, bem organizado e bem redigido. A divisão de tarefas do trabalho se aplica somente à implementação. Todo o grupo é responsável pelo relatório (escrita e revisão).

5. Exigências

O trabalho deverá, obrigatoriamente, atender aos seguintes requisitos:

- Implementação em C ou C++
- O projeto deve ser compilável e executável via linha de comando. Não conte com a presença de IDEs como Code::Blocks, Visual Studio ou NetBeans. Caso seu grupo opte por utilizar algum ambiente de desenvolvimento, certifique-se de que o projeto enviado possa também ser facilmente compilado em um sistema operacional Linux sem esses ambientes instalados. Forneça instruções claras e precisas de compilação e execução pela linha de comando. **Recomenda-se a utilização de algum Makefile ou script para a compilação.** Caso o grupo julgue necessário, é possível solicitar que o professor verifique as instruções de compilação **antes** do prazo final de envio.
- **Trabalho entregues após o prazo serão penalizados**

- O programa desenvolvido deve permitir que o usuário entre com o caminho do diretório que contém os arquivos de dados como um argumento na linha de comando. Veja o Exemplo:

```
$ ./programa /diretorio/contendo/arquivos/entrada
```

- O programa deve procurar pela existência dos arquivos binários dentro da pasta. Se não existir, deve pré-processá-los (etapa 1) antes de seguir para o módulo de testes (etapa 2).
- Não é permitida a utilização de bibliotecas externas. O processamento do arquivo CSV e os todos os algoritmos devem ser implementados pelo grupo. Um dos objetivos do trabalho é que vocês aprendam a trabalhar com arquivos e com os algoritmos escolhidos. Se você tiver dúvida quanto à utilização de alguma função ou biblioteca, entre em contato com o professor.
- Obviamente, todo código deve ser de autoria do grupo. Não é permitida a utilização de códigos de terceiros ou de outros grupos. É permitida a pesquisa por estratégias para a solução dos problemas (e as referências utilizadas nessas pesquisas devem constar do relatório), porém a apropriação de código alheio não será aceita. **Qualquer tentativa de plágio identificada resultará em nota zero. Os códigos fontes serão analisados pelo sistema Moss** (<http://theory.stanford.edu/~aiken/moss/>)

6. Entrega

O grupo deverá ser formado por **no máximo** 4 alunos, e as responsabilidades de cada aluno devem ser documentadas e registradas. Não é permitido que algum integrante do grupo fique responsável somente pela confecção do relatório. Todos os integrantes devem contribuir com a implementação. A distribuição das responsabilidades deve ser feita de maneira uniforme, de modo que cada membro do grupo se envolva com o trabalho na mesma proporção que os demais.

Todos os itens abaixo devem ser entregues:

1. Código-fonte completo;
 - a. Deve ser submetido um link para um repositório git (github) contendo o código do trabalho;
 - b. **Não incluir o dataset na submissão via Google Classroom (tanto o .csv quando .bin).**
2. Relatório em Google Doc atendendo ao especificado na Seção 4 deste documento.

7. Critérios de avaliação

O grupo será avaliado de acordo com os seguintes critérios:

- Execução correta do código (E); (0-100)
- Atendimento ao que foi solicitado (A) (valor entre 0 e 1);
- Organização do código (O): seu código deve estar bem modularizado e bem documentado; (0-100)
- Qualidade do relatório apresentado (R). (0-100)

Cada membro do grupo será avaliado individualmente, tanto com relação aos detalhes de implementação que ficaram sob sua responsabilidade, quanto ao entendimento em alto nível de abstração do que foi feito pelo grupo como um todo. O entendimento teórico do conteúdo relacionado ao trabalho também será avaliado. A nota individual (M) será um valor de 0 a 1 que irá ponderar a nota da implementação.

A nota final de cada integrante será computada de acordo com a seguinte fórmula:

$$\text{Nota} = 0.6 * [(0.7 * E * A + 0.3 * O) * M + 0.4 * R * A]$$

IMPORTANTE:

Caso algum membro do seu grupo tranque ou abandone a disciplina, comunique o professor o quanto antes para que se possa discutir alternativas.

Referências

[1] Yamac Eren Ay. Spotify DataSet 1921-2021 (2021). Disponível em: <https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks> (acessado em 25 de maio de 2021).