

Projeto Laboratórios de Informática III - Fase 1

Projeto elaborado por:

Ricardo Sousa (A104524)
Tomás Oliveira (A104345)
Pedro Argainha (A104351)

Grupo 62

Licenciatura em Engenharia Informática

Departamento de Informática, Universidade do Minho

Introdução

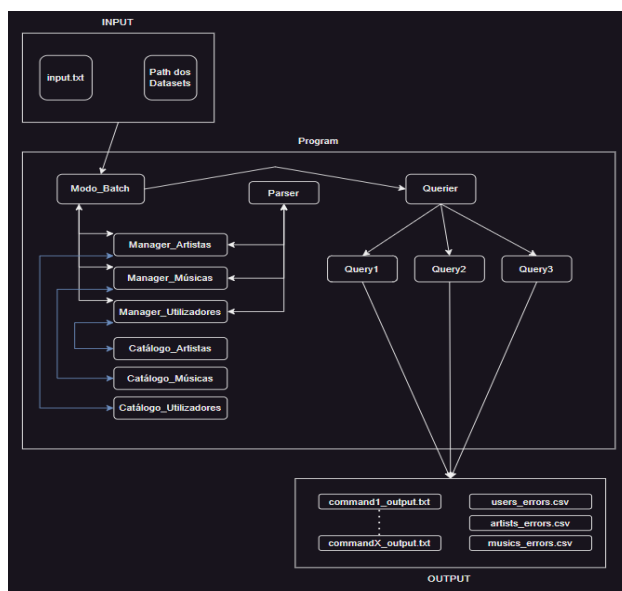
Este projeto está em desenvolvimento como parte da unidade curricular de Laboratórios de Informática III no ano letivo 2024/2025. A unidade tem como objetivo apresentar aos alunos os princípios fundamentais da Engenharia de Software, como modularidade, reutilização, encapsulamento e abstração de dados. A tarefa consiste em implementar uma base de dados em memória para armazenar dados fornecidos pelos docentes.

O projeto está dividido em duas fases, sendo que a primeira fase concentra-se na implementação do parsing dos dados e no modo batch. Neste modo, várias queries são executadas sequencialmente por dados armazenados num arquivo de texto, cujo caminho é passado como argumento. Um exemplo de uma query é "Listar o resumo de um utilizador, consoante o identificador recebido por argumento". Este processo envolve a leitura, interpretação e armazenamento eficiente dos dados em estruturas de dados versáteis. A utilização da biblioteca glib facilita a implementação dessas estruturas.

Desenvolvimento

Dada a complexidade do projeto, o grupo realizou uma extensa fase de planeamento antes de iniciar a codificação. A arquitetura proposta pelos docentes serviu como base, e foi crucial pensar na comunicação entre os diversos módulos, garantindo a presença de encapsulamento e modularidade em todo o projeto.

A arquitetura resultante assemelha-se à idealizada pelos docentes:



A escolha das estruturas ideais para armazenar os dados foi um desafio inicial. Optou-se por hashtables para as estruturas principais, dada sua eficiência no acesso. Três hashtables foram criadas para usuários, artistas e músicas, agrupadas numa estrutura chamada de catálogo, seguindo a arquitetura mencionada. O parsing dos dados também foi considerado. Criou-se um manager para cada entidade e cada um manda para o parser o ficheiro .CSV correspondente onde este devolve o número de linhas lidas e as linhas em si. O manager de seguida, valida cada linha e, caso algum dado de uma linha seja inválido, essa linha é guardada num ficheiro de erros. Caso contrário, essa linha é guardada na hashtable.

Para concluir a primeira fase, foram implementadas as queries 1, 2 e 3. A explicação detalhada do processo de implementação é apresentada a seguinte:

Query1: Listar o resumo de um utilizador, consoante o identificador recebido por argumento. Dado o identificador, é de seguida recolhida informações específicas sobre o utilizador, sendo a data de nascimento o único dado que necessita de ser processado para obter a idade do utilizador. Esses dados todos são apresentados de acordo com a estrutura dada no enunciado.

Query2: Apresentar os top N artistas com maior discografia. Esta query primeiro verifica se é apresentado o argumento limitador de país. Se sim, esta percorre a lista de artistas e sempre que passa num com o país certo percorre a lista de músicas somando as discografias. Aí, ele guarda os artistas num array ordenado de N posições. Quando acaba escreve no ficheiro os artistas do array.

Caso não haja um limitador de país, a query itera pelas músicas todas adicionando os artistas e a sua discografia a uma lista ligada. No fim, ordena e procura os top N artistas escrevendo-os no ficheiro

Query3: : Apresentar os géneros de música mais populares numa determinada faixa etária. Dada a faixa etária, é de seguida percorrida a hashtable de utilizadores e caso este esteja dentro da faixa etária, verifica-se a sua lista de gostos. É verificada qual o género de cada música e incrementa-se o contador do género correspondente. No final teremos cada contador com o número de gostos de todos os utilizadores da faixa etária. De seguida, é escrito no ficheiro de output esses resultados de forma decrescente.

Dificuldades Encontradas

O projeto apresentou desafios significativos, especialmente na implementação cuidadosa do encapsulamento e modularidade. Reconhece-se que esses princípios são fundamentais para um programa de qualidade, e a primeira abordagem pode conter falhas que precisam de ser melhoradas na segunda fase. Outra dificuldade encontrada foi na redução do tempo de execução que, embora o tenhamos conseguido diminuir por uns segundos, ainda é preciso mais alguma otimização para a segunda fase.

Conclusão

Em resumo, o projeto provou ser desafiador para o grupo, exigindo estratégias aprimoradas para aperfeiçoar o trabalho atual e prepará-lo para a avaliação na segunda fase.