

# **Instituto Federal de Educação, Ciência e Tecnologia do Norte de Minas Gerais - IFNMG**

## **Curso de Bacharelado em Ciência da Computação**

### **Organização e Sistemas de Arquivos**

Prof. Wagner Ferreira de Barros.

Email.: [wagner.barros@ifnmg.edu.br](mailto:wagner.barros@ifnmg.edu.br)

Prazo: 2 semanas (Final: 28/03 – Domingo)

### **Trabalho Prático 2 – Parte 2 – (Dupla: a mesma da parte1)**

#### **Indexação - Programas de TV e filmes listados no Netflix**

O conjunto de dados disponibilizado, consiste em programas de TV e filmes disponíveis no Netflix a partir de 2019. O conjunto de dados é coletado do Flixable, um mecanismo de pesquisa de terceiros da Netflix.

Em 2018, eles lançaram um relatório interessante que mostra que o número de programas de TV na Netflix quase triplicou desde 2010. O número de filmes do serviço de streaming diminuiu em mais de 2.000 títulos desde 2010, enquanto seu número de programas de TV quase triplicou. Será interessante explorar o que todos os outros insights podem ser obtidos no mesmo conjunto de dados.

Integrar este conjunto de dados com outros conjuntos de dados externos, como classificações “IMDB”, “rotten tomatoes” também pode fornecer muitas descobertas interessantes.

#### **Dados consideradores no arquivo (obtidos da descrição original do arquivo):**

- show\_id (texto): *Unique ID for every Movie / Tv Show*
- type (texto): *Identifier - A Movie or TV Show*
- title (texto): *Title of the Movie / Tv Show*
- director (texto): *Director of the Movie*
- cast (lista de texto, separado por vírgulas): *Actors involved in the movie / show*
- country (texto): *Country where the movie / show was produced*
- date\_added (data -> dd/mm/aaaa): *Date it was added on Netflix*
- release\_year (short): *Actual Release year of the move / show*
- rating (texto, há 5 tipos diferentes): *TV Rating of the movie / show*
- duration (texto): *Total Duration - in minutes or number of seasons*
- listed\_in (lista de categorias, separadas por vírgulas): *Genere*
- description (texto): *The summary description*

(abra o arquivo em um editor de texto, ou importe em uma planilha eletrônico para poder entender melhor o conteúdo de cada campo)

#### **Objetivo:**

Neste trabalho prático vocês deverão implementar um sistema de indexação que permita recuperar de forma rápida informações para algumas pesquisas específicas descritas abaixo:

- 1) gerar 1 arquivo de índices direto mapeando o ID de todos os filmes e séries do arquivo.
- 2) gerar 1 arquivo de índice indireto para recuperar o id de um filme ou série, pelo seu título.
- 3) gerar 2 arquivos de índices mapeando os ID's de todos os filmes e de todas as séries.
- 4) gerar um arquivo de índice para todos os filmes e séries brasileiras.
- 5) implementar um método que permita pesquisar por um filme ou série pelo seu título ou parte dele (neste caso, exiba todos os registros que contenham a parte do título desejado).

- 6) permitir a exibição de todos os filmes brasileiros lançados no ano de 2019.

### **Diretrizes de Implementação**

- a) Todo o seu código deverá estar em C++, além de apresentar boa modularização (uso de classes e arquivos separados coerentes com cada classe), endentação e comentários (que não poluam o código, mas permitam uma boa leitura e manutenção).
- b) Cada arquivo de indexação deverá ser formado por registros de apenas dois campos (“chave pesquisa”, “endereço do registro referente a chave”).
- c) Deverá ser criada uma classe para manipulação do registro a ser escrito/lido no arquivo. Aconselha-se também a criação de uma classe capaz de realizar o acoplamento dos registros com o arquivo e vice-versa (tendo como atributos os streams de entrada e saída, e dados auxiliares, por exemplo), método de leitura do arquivo de entrada, geração de índices, realização das buscas, etc..
- d) O formato do arquivo de entrada (netflix\_titles.csv) não poderá ser alterado.
- e) Os arquivos de índices deverão ser gerados imediatamente após a leitura do arquivo de entrada (com formato de livre escolha, mas aconselha-se o uso do formato binário).
- f) Deverá ser fornecida uma tela de interação do usuário para permitir que este acesse as funções de pesquisas e exibição de dados implementados.