

UNIVERSIDADE DE BRASÍLIA

Faculdade do Gama

Sistemas de Banco de Dados 2

Trabalho Final (TF)

Visualização - F11

Thiago Cerqueira Borges - 190117508
Guilherme Soares Rocha - 211039789

Brasília, DF
2023

Base de Dados (documentação e motivos)

My Anime List ou MAL é um site e aplicativo popular dedicado à comunidade de fãs de anime e mangá. Ele funciona como uma base de dados abrangente para anime, mangá e light novels, permitindo que os usuários pesquisem, classifiquem, avaliem e revisem uma ampla variedade de obras relacionadas à animação japonesa e à cultura otaku em geral. A base de dados escolhida contém informações sobre 17.562 animes e a preferência de 325.772 usuários diferentes.

Após uma longa discussão a base foi escolhida pois os dois integrantes tem familiaridade com tema, então escolhemos anime-recommendation-database-2020 base recomendada pelo aluno Thiago Borges, ao invés da sugerida pelo Guilherme com o tema de finanças, a base se encontra na plataforma Kaggle.

O banco de dados pode ser encontrado em:

<https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020?select=watching_status.csv>

Dificuldades e Problemas

Ao realizar a importação do banco de dados disponível no Kaggle, que estava no formato CSV, para o MySQL, deparei-me com desafios significativos. A principal dificuldade surgiu da presença de caracteres japoneses no conjunto de dados, os quais não foram processados adequadamente durante o processo de importação. A codificação dos caracteres, juntamente com as peculiaridades da língua japonesa, resultou em entradas incorretas e dificuldades de interpretação dos dados. Além disso, constatei que o banco de dados não estava estruturado de acordo com a Primeira Forma Normal (1FN), um dos princípios fundamentais de normalização, no caso o atributo Genres possui mais de uma classificação o tornando divisível. Esta constatação acrescentou uma camada adicional de complexidade à integração dos dados, exigindo ajustes e modificações na estrutura do banco de dados para garantir a conformidade com as regras da 1FN, também não possui chaves primárias. Este processo revelou-se desafiador, mas valioso para aprimorar a qualidade e a integridade dos dados armazenados no MySQL.

Após as aulas de SBD2 é notório que a escolha da base de dados não foi boa, se tivéssemos o conhecimento adquirido nas aulas não teríamos escolhido essa base.

Diagrama Entidade Relacionamento - DER

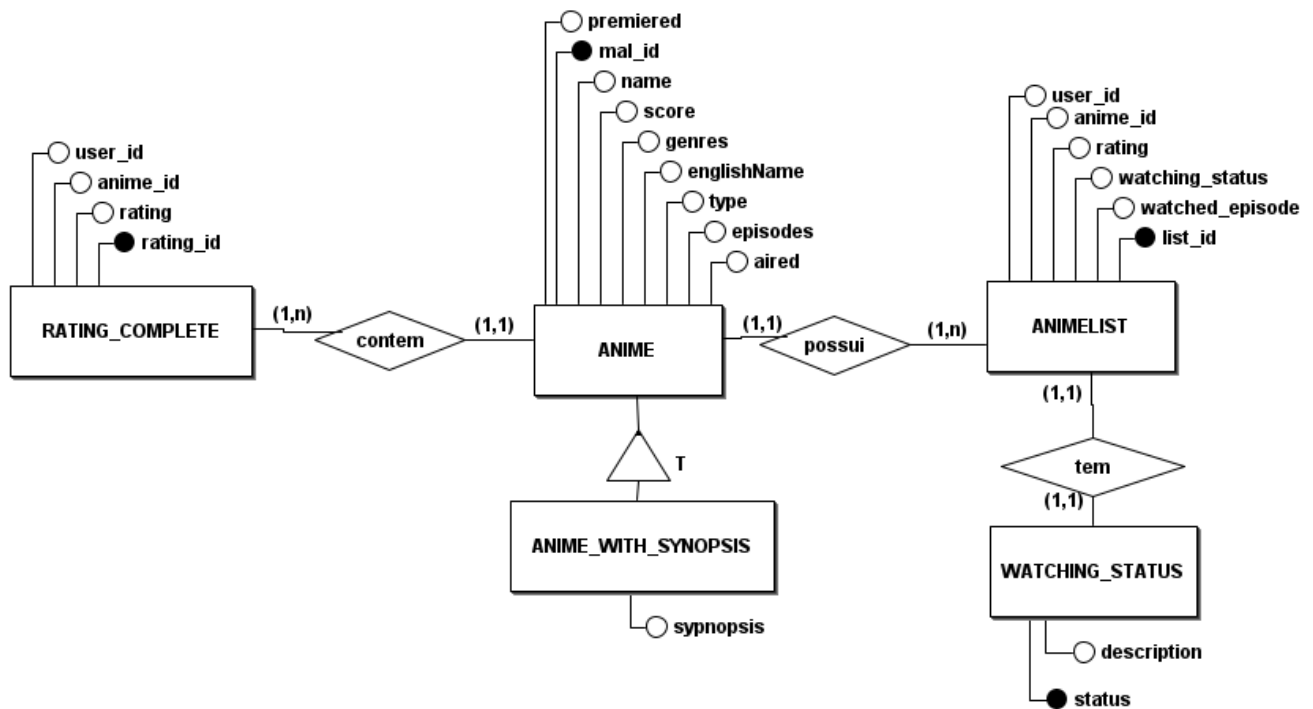
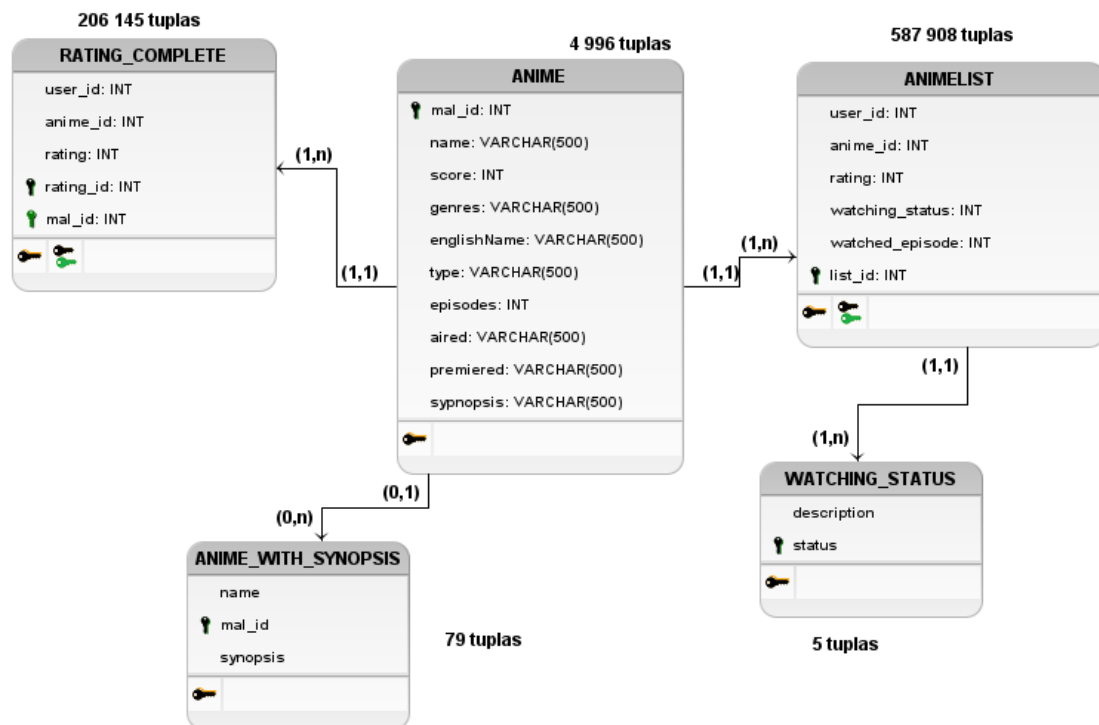


Diagrama Lógico de dados - DDL



Dicionário De Dados

Entidade: ANIME				
Descrição: Informações do anime retiradas da página principal e da página de estatísticas do site My Anime List.				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
mal_id	chave primária obrigatório	int	3	ID MyAnimelist do anime.
name	obrigatório	varchar	255	Nome do Anime
score	obrigatório	float	3	pontuação média do anime dada por todos os usuários no banco de dados MyAnimelist.
genres	obrigatório	varchar	255	lista separada por vírgulas de gêneros para este anime.
englishName	optativo	varchar	255	nome completo em inglês do anime.
type	optativo	varchar	45	TV, filme, OVA, etc.
episodes	optativo	int	3	número de capítulos.
aired	optativo	varchar	100	data de transmissão.
premiered	optativo	varchar	45	estreia da temporada.

Entidade: ANIME_WITH_SYNOPSIS				
Descrição: Informações do anime retiradas da página principal e da página de estatísticas do site My Anime List, porém com sinopse.				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição

mal_id	chave primária obrigatório	int	3	ID MyAnimelist do anime.
name	obrigatório	varchar	255	Nome do Anime
score	obrigatório	float	3	pontuação média do anime dada por todos os usuários no banco de dados MyAnimelist.
genero	obrigatório	varchar	255	lista separada por vírgulas de gêneros para este anime.
englishName	optativo	varchar	255	nome completo em inglês do anime.
japaneseName	optativo	varchar	255	nome completo em japonês do anime
type	optativo	varchar	45	TV, filme, OVA, etc.
episodes	optativo	int	3	número de capítulos.
aired	optativo	varchar	100	data de transmissão.
premiered	optativo	varchar	45	estreia da temporada.
synopsis	optativo	varchar	255	breve resumo (sinopse) do anime.

Atributo multivalorado: **RATING_COMPLETE**

Descrição: contém a lista de classificações dadas pelo usuário aos animes.

Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
----------	--------------------------	--------------	---------	-----------

user_id	chave primária obrigatório	int	3	Identificador do usuário
anime_id	chave estrangeira obrigatório	int	3	identificador do MYANIMELIST do anime que este usuário avaliou.

Atributo multivalorado: WATCHING_STATUS				
Descrição: Descrição de todos os status possíveis da coluna: "watching_status" em animelist				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
status	obrigatório	int	3	Identificador do status
description	obrigatório	varchar	255	descrição do status

O banco de dados pode ser encontrado em:

<https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020?select=watching_status.csv>