Dataset Preparation
- Anime List -

DAPI

Master in Informatics and Computing Engineering

2018/2019

Andreia Rodrigues - 201404691
Francisco Queirós - 201404326
Miriam Gonçalves - 201403441

U. PORTO
FEUP **FACULDADE DE ENGENHARIA**
UNIVERSIDADE DO PORTO

Introduction

Anime ⟶ Animation

➜ Japanese-style animated film or tv show.

➜ Characterised by the characters having huge eyes, bright colored hair and exaggerated emotional expressions and gestures.

➜ Originate from manga, visual novels, light novels or video game adaptations.

➜ Popular form of entertainment for adults as well as for younger audiences.

Introduction

➜ Anime fans gather in online platforms where **information about animes and anime reviews are collected** and can be accessed.

➜ These online platforms **gather their own data separately** and **there isn't a centralized place** where users can have an overview of an anime and the rating it is given in each of them.

Information Retrieval

kaggle
MyAnimeList

\+

IMDb

\+

🦊 KITSU

Used to get general information about animes and their rating in *MyAnimeList*.

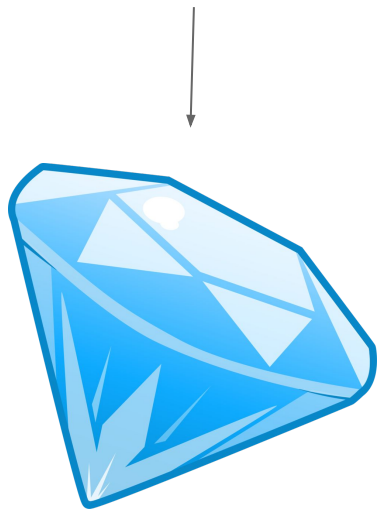Used to get rating information from IMDb.

Used to get additional information about animes and their rating in Kitsu.

Data Preparation
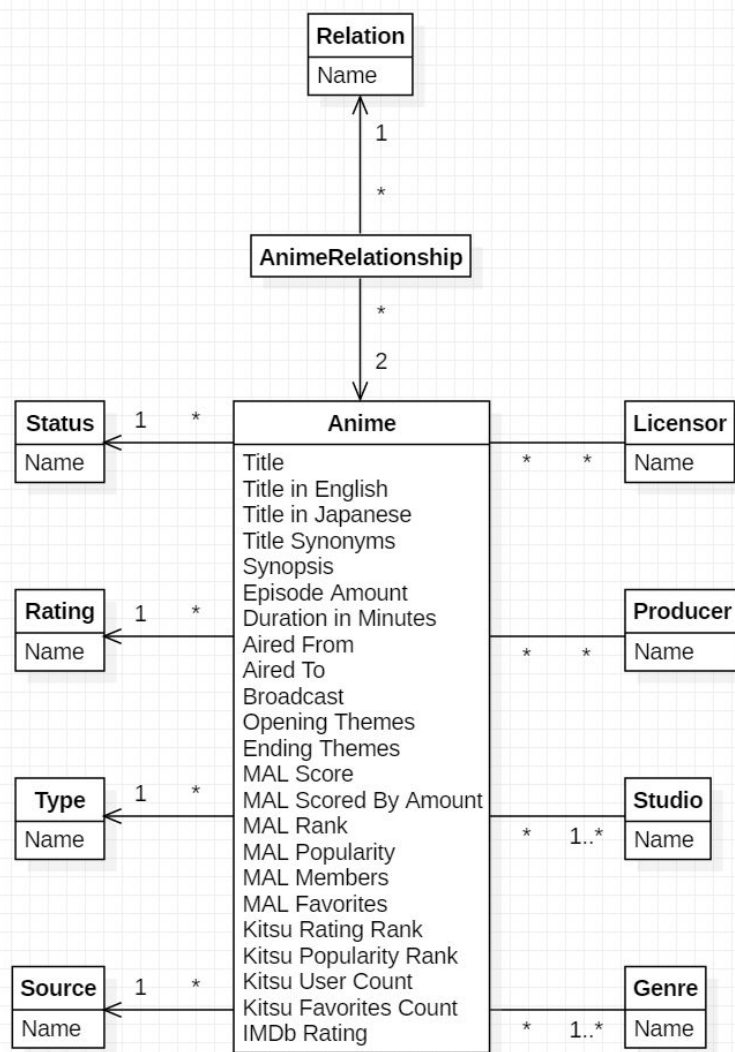
1st step

Cleaning and organizing the dataset

2nd step

Enrich the data

Data Model

Data Limitations

▷ Crossing data between different datasets

▷ Data updates