Project - Phase 8 Report

Group 14
Tiago Carvalho fc51034
Diogo Lopes fc51058
João Roque fc51080
Miguel Saldanha fc51072
João Afonso fc51111

16/05/2021

1 Motivation

For our project we were thinking about an API that could help us decide which shows to watch next, by marking shows as viewed and/or liked.

Because we can watch more than just movies, like anime, we wanted to use more than one dataset. Both animes and movies can have a lot in common not only with each other but also with books, so we also decided to use a book dataset.

So we have three datasets, and we can effortlessly search thought any of them, mark them as seen or liked, and get suggestions. Having very similar categories in every single one of them.

For the suggestions our idea was making a recommendation list having in mind the item's rating and user's likes and views, which would indicate to us which categories the user prefers.

So it makes sense to call our API "Seen".

2 Dataset characterization

2.1 Dataset 1 — IMDB

This data set provides a lot of information about movies and shows that can be seen in IMDB.

We downloaded the dataset from the Kaggle website, updated one year ago.

From the whole data this where the columns that were important to us:

Columns	Example
id	606e2683b3fff1da8a207ae9
name	The Arrival of a Train
category	[Action,Documentary,Short]
rating	7.4
type	short

2.2 Dataset 2 — MyAnimeList

For the second data set we got it from Kaggle, about the MyAnimeList website.

This data not only has a lot of anime content but also user information, but because we want to connect with the other datasets doesn't make sense to use that data. Meaning we used these columns:

Columns	Example
id	606e252aebddc73ebfb15507
name	Shakugan no Shana: Season II
category	[Action,Drama,Fantasy,Romance,School,Supernatural]
rating	7.72
imageUrl	https://myanimelist.cdn-dena.com/images/anime/10/18669.jpg

2.3 Dataset 3 — GoodReads

At last, this data set represents books from the GoodReads website, also downloaded from Kaggle.

The helpful data from this data set, to be able to use with animes and movies, is its categories and rating:

Columns	Example
id	606e25ad5e927a606f534284
name	Of Mice and Men
description	The compelling story of two outsiders [...]
category	[Classics,Fiction,Academic,School,Literature,Historical]
rating	7.7
imageUrl	https://images.gr-assets.com/books/1511302904l/890.jpg

3 Use cases

We have 3 types of Users: an Admin, which is a logged-in user with special permissions, a Regular user, which is a logged-in user, and a not logged-in user that we call Any.

Services	User	Functionalities
Normal	Any	Sign in See Book, Show and Movie Library
	Regular	User Log in Set Book/Show/Movie as seen Set Book/Show/Movie as liked Ask for suggestions to read and/or watch
	Admin	Add Book/Show/Movie to Library Remove Book/Show/Movie from Library
Spark	Any	Count how many views a specific Item has Count how many likes a specific Item has Top 10 Items with more likes

4 API

User	Path				get	post	put	del	description
Regular	/lib	/{page}			X				Returns a *page* from the database
	/suggest					X			List of suggestions to watch
Admin	/item					X			Creates an item to add to the database
Any	/item	/{type}	/{id}		X			X	Gets/Deletes item with specific *id* and *type*
Regular	/item	/{type}	/{id}	/seen			X		Marks item as seen
	/item	/{type}	/{id}	/like			X		Marks item as liked
Any	/item	/{type}	/{id}	/views*	X				Returns Item's number of views
	/item	/{type}	/{id}	/likes*	X				Returns Item's number of likes
	/getTopTen	/{type}			X				Returns top ten most liked Items with *type*
	/user					X			Creates User
	/user	/login			X				Logs in
Regular	/user	/logout			X				Logs out
	/user	/search	/{username}		X			X	Searches/Deletes User by username

5 Architecture (application and technical)

5.1 Diagram

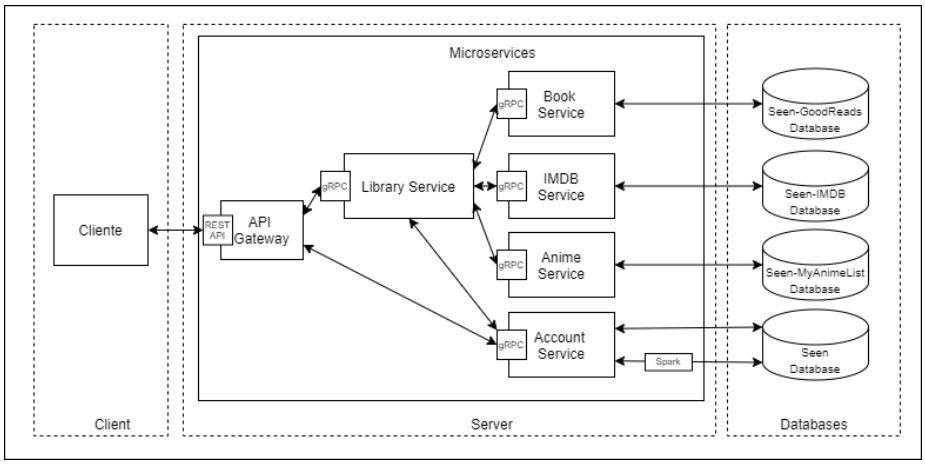

Figure 1: TODO

6 Implementation

7 Evaluation and validation

8 Cost analysis

9 Discussion

10 Conclusions