TTDS Group Project

Movie Index - Search Engine for Internet Movie Database

S2322928, S1837774, S2449890, S2402728, S2443454

Mar 2023

# Abstract

# Introduction

Movies, TV shows, anime has become the most common means for people to get relaxed and seek pleasure while enjoy the the leisure time. While the trend of these films and TV shows dominating the large portion of daily entertainment market, it could require some efforts to find the desired films which people wish to watch and the result could be fairly frustrating. Thus, a more advanced search engine targeting on these information can be significantly useful in the real-life scenario.

This project, therefore, was designed to address the problem of struggling in finding films to watch by providing a more progressive solution which embedded a bunch of carefully selected feature.

In comparison to the search function offered by [IMDB.com](http://IMDB.com)*,* which is only able to query the title of films, this search engine dives deeper into the relevant information available for any film such as storyline, genre, keywords to present users more possibilities when they are trying to find a film to enjoy. Essentially, the concept behind this is to explore every single attribute of a film and presenting the result obtained by ranking algorithms, which will be described in later sections. Furthermore, it also provides advanced search to fulfil the real-life scenario needs - such as searching by keywords and genre, filtering the result by colour information, year of releasing, etc.

This system has embedded with nearly 115,000 documents in XML format, being a subset of the Internet Movie Database (IMDB), which was available to be downloaded from https://data.mendeley.com/datasets/rth2kr5hxf/1. Additionally, in order to make the system more scalable in the future, a fully-functional web scraping class was integrated into the system so that more up-to-date data can be effortlessly appended to the existing dataset from time to time.

Regarding the development process, this project was built by a team of 5 students from different disciplines and the strictly followed the concepts of Agile software development life cycle. As the sprint was weekly, all the formal meetings held and meeting minutes file was documented in order to track the progress periodically. While the UML design of the project was unfortunately lost, the mock-up of the original front-end design was provided alongside with the code in the *Design* folder.

Upon the successful completion of the most basic function of the search engine, more advanced features such as auto-translation detection, query expansion, searching suggestion, relevant movie redirection, etc was implemented and will be discussed more in-depth in the section of ***Additional Features***.

# System Design

# Data Collection and Storage

# Preprocessing

# Indexing

# Ranking of Results

# Advanced Search

# API

# GUI

# Evaluation

# Development Life Cycle & Individual Contribution

# Additional Features

# Documentation