

TTDS Group Project Plan

Team 24

January 28, 2021

1 System functions

1.1 Compulsory functions

We intend to create a Movie Reviews IR System which can provide the following features:

1. User can search for specific movie reviews by queries(Free text, Boolean and phrase/proximity search) (CW1 content).
2. Find the top tokens(Mutual Information or X2 scores or TFIDF) of reviews for the same movie and show them as keywords of reviews for the movie(cw2 Token Analysis).
3. User can search a movie name and get all the reviews for the movie.(How results will be displayed: headings of reviews) The sequence of results can be sorted by review time or review score.
4. Each movie has genres, user can search for a genre and the system can show all the movies belong to it. The sequence of results can be sorted by movie rating, publish time.
5. User can search a movie name and get information for the movie including: actors, director, publish date...
6. This is a live data collection system. User can write review and decide if a review is "helpful" or not.(Without registering an account)
7. Multi user can access the system at a time.

1.2 Optional functions

Functions are considered to add if everything goes well:

1. Query suggestion / Spell checker
2. IR evaluation (CW2 content).
3. Query Expansion (Display learnt terms with search)
4. Account management system (User can register account)

2 User Interface

This system can be accessed on desktop browsers (movable devices interface won't be available).

3 Data Collections

Million records, English only, text only

1. IMDb Largest Review Dataset part-1 (Number of reviews = 1, 010, 293)
[1]
(<https://www.kaggle.com/ebiswas/imdb-review-dataset>)
2. MovieLens 25M Dataset [2]
(<https://grouplens.org/datasets/movielens/25m/>)

4 Server

Tencent Cloud (<https://intl.cloud.tencent.com/>)

References

- [1] Enam Biswas. Imdb review dataset, 2021.
- [2] F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4), December 2015.