



MovieLike

CS 410 Final Project

Shubham Jain (shubham9), Sameksha Reddy (screddy2)



Project idea

- Crawler: Crawls over imdb website to get movies and their details like poster, rating etc and their reviews
- Search Engine: We add a search engine to search for movies based on their reviews to efficiently match user preferences to opinions from reviews.



Repository

Code is available here : <https://github.com/shubham1310/cs410project>



Crawler - part 1

Gets the movie names, their rating, genres, and their imdb id

Link :

<https://github.com/shubham1310/cs410project/tree/master/Crawler>



Crawler - part 2

Get the movie review, poster, plot etc.

Link :

<https://github.com/shubham1310/cs410project/blob/master/crawler.ipynb>



Search Engine (Opinion based Ranking)

- Vast amount of opinions (movie reviews from IMDb)
- Rank entities given user preferences (good climax)
- Support for multiple preferences



Process

- Entities -> movies represented by reviews
- BM25 as default retrieval function
- Uni-gram and bi-gram analyzers
- Query Aspect Modelling
 - Handle multiple preferences
 - Geometric mean of the scores to rank
- Opinion Expansion
 - Handle non uniformity in expressing multiple movies (good movie, great movie)
 - Expand common praise words



Demo

To install and do things from scratch