Proposal

Haotian Qiao 31695920

My project will focus on movies. I am going to use both Reddit and IMDb (Internet Movie Database). IMDb provides plenty of interfaces with respect to movies, such as searching by titles and genres ,top movies and box office. Reddit is a popular forum where people with the same interests can discuss various topics. On Reddit, there is a group called "r/ movies"and people discuss and share their opinions on different movies.

The whole idea is that users can search movies by titles or genres to get information from IMDb. The results contain basic information (including posters url, wiki url, Youtube url and so on) about this movie. Then, people can search it on Reddit to view the comments. Also, if users cannot figure out the specific movies, the api of IMDb could return movies of some genres. After having the return results, users can find their interested movies so that they can search them on Reddit to see if they really want to watch them.

I picked some apis from IMDb and I will focus on them. I will cache the searching history (not all returned results but the items users clicked) both for IMDb and Reddit. Also, to have some statistics of searching history, I will cache the numbers for each genre in history, the total searching times through the whole history for IMDb and Reddit and the numbers of each rating range. I plot two line charts and two bar graphs.

For the tree part, I build a decision binary tree to guide the users. It will return the "movielist" depending on the user's input. For example, if users want to search with genres, they will eventually input the genres and get the result list.