507 Final Project Proposal:

Chuhan Hou

Section 8

Project overview:

A program that first scrape information from MIT the media lab research site to grasp a list of projects and general overview of the projects. Then the most frequent keyword would will be used to search for movies that won prizes for the past 10 years.

then give user options to see more analytics about the movies’ directors.

Data sources:

Get and cache data from Open movie database(4)

Scraping information from MIT the media lab. (4)

Presentation options:

Get information about movies that won best pictures for the past 10 years along with their related information (release data, director, play write, actors etc)

Rankings of directors who won awards for the past 10 years.

Presentation tool:

Plotly and interactive queries.

Tables:

1. Labs table (Group)
   1. Lab Id
   2. Lab name
   3. Lab brief
   4. Project Instructor
   5. PI Email
   6. Number of people in a lab
   7. Most recent updates (year)
2. Projects in the labs
   1. Project ID
   2. Lab Id
   3. A. project name
   4. Project brief
   5. Hashtag number
3. Hashtag table
   1. Id
   2. Name
4. Junction table of hashtag and project
   1. Hashtag name
   2. Project ID

Analysis:

1. Top 5 non-stop word in the project brief
2. Top 5 non-stop word in lab brief
3. Top 10 hashtags
   1. Histograms of hashtags in each lab group
4. Histogram of number of people in each lab
5. Potentially wordnet

All the research group available in ML:

Pick one for brief, PI Info and num of people, most recent update, top projects

Choose one project for more info:

Associated hashtags

Visualize the people size distribution across labs:

Visualize the hashtags