

ASSIGNMENT 6

- Anirudh Narayanan(an9425)

For all questions. Set proper path to the mongodb collection.

Pre-process data:

Run q2_cleanup.py

Enter the path of the json_data provided for the assignment

Edit the `json_data_cleaned_final.json` to add '[' to the starting position and remove ',' and add ']' to the last position of the file.

Create Mongodb Database using the following command line terminals

1. % mongoimport --db movies --collection movies --file
"/Users/anirudhramesh/Desktop/INTRO_TO_BIGDATA/ASSN4/movies.json" --drop
2. % mongoimport --db movies --collection new_data --file
"/Users/anirudhramesh/Desktop/INTRO_TO_BIGDATA/ASSN4/json_data_cleaned_final.json"
--drop --jsonArray
3. % mongoimport --db movies --collection movies_backup --file
"/Users/anirudhramesh/Desktop/INTRO_TO_BIGDATA/ASSN4/movies.json" --drop

Q2:

Run q2.py

Q3

Run q3.py

Commented part of the code uses the logic of storing the labels visited and skipping them and updating only 1 document per match instead of many.

Q4

1. Run Q41.py
2. Run Q42.py
3. Run Q43.py