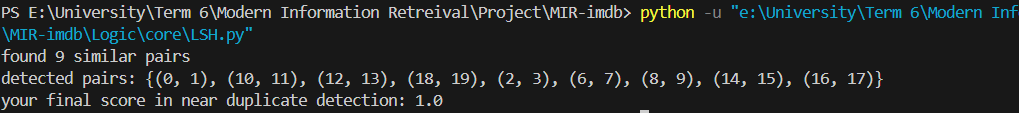
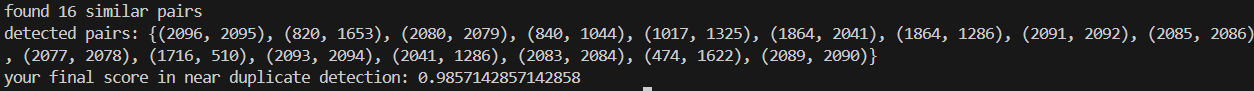
# LSH Correctness:

We first test on only fake data and see the models performance:



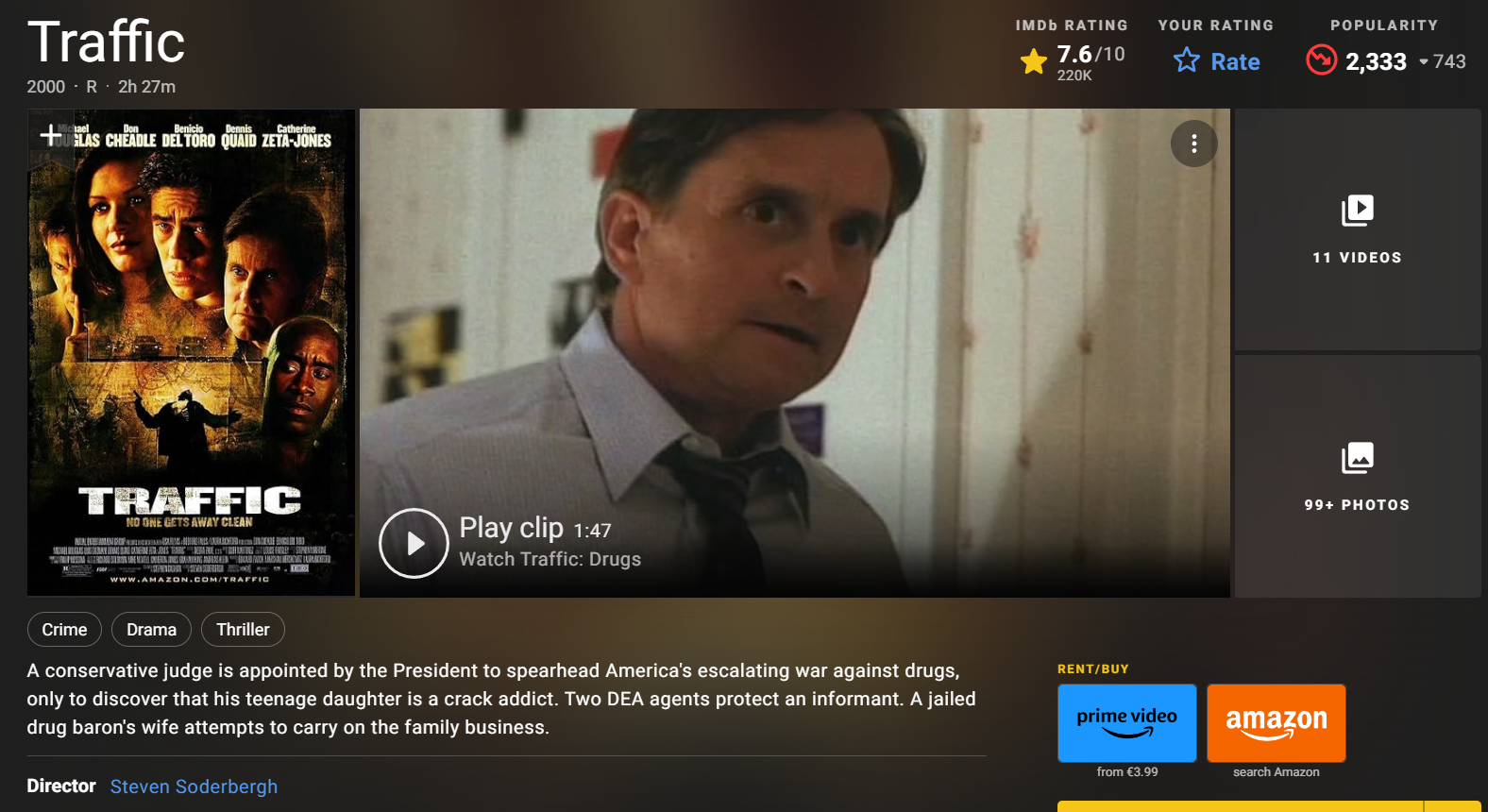
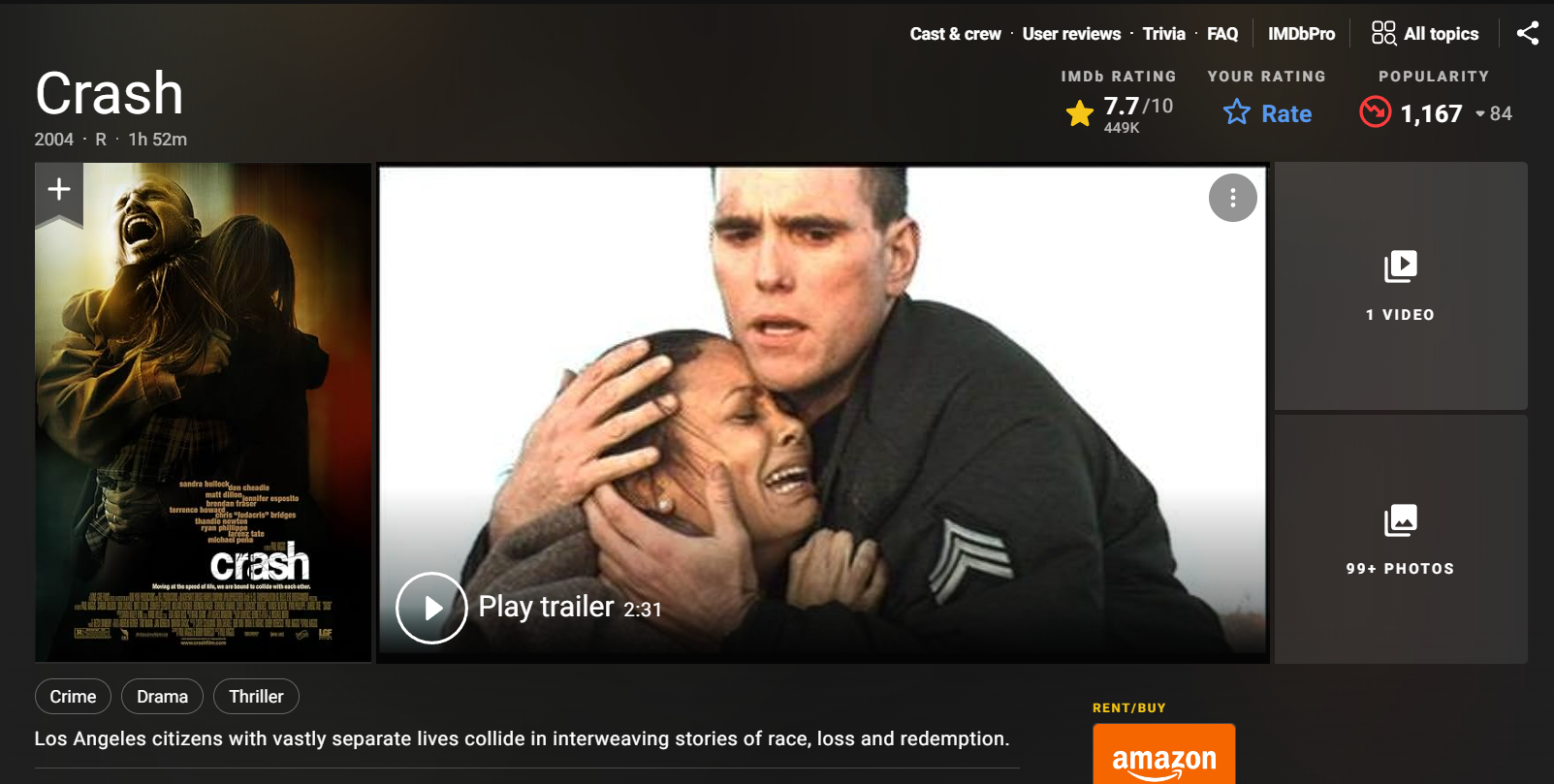
As you can see model detected 9 out of 10 similar movies correctly. Also the score is 1

Now let’s test on all the movies in corpus:



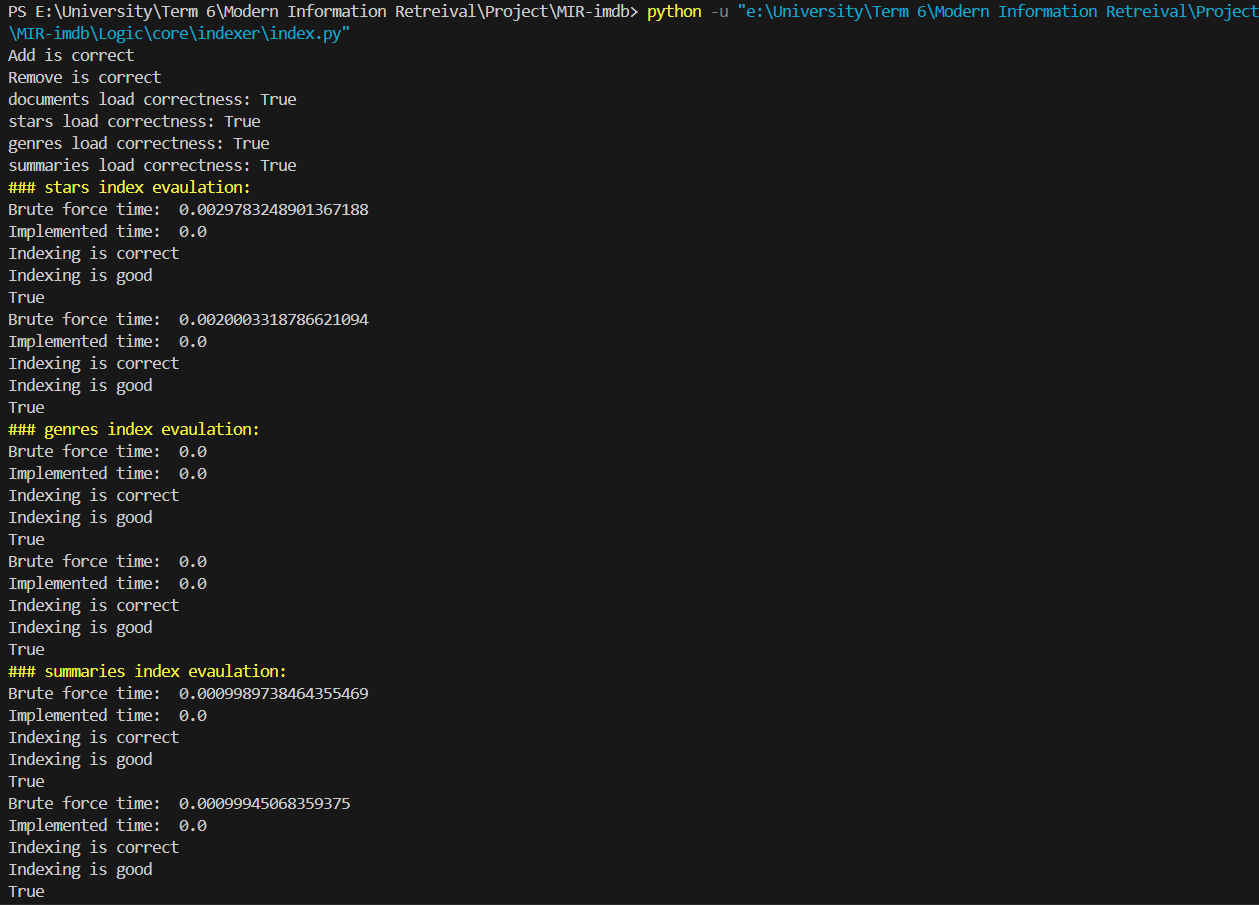
As you see the score is still very high. And some extra pairs have been detected from our data as well.

Example of a pair of movies detected as similar:



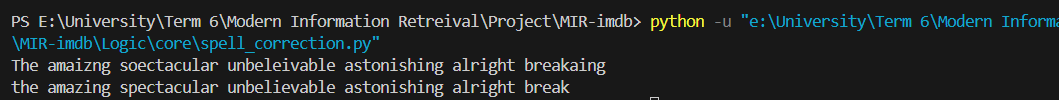
# Index correctness:

We perform several test to check the correctness of index and compare it’s performance against the brute force method.



# Spell correction:

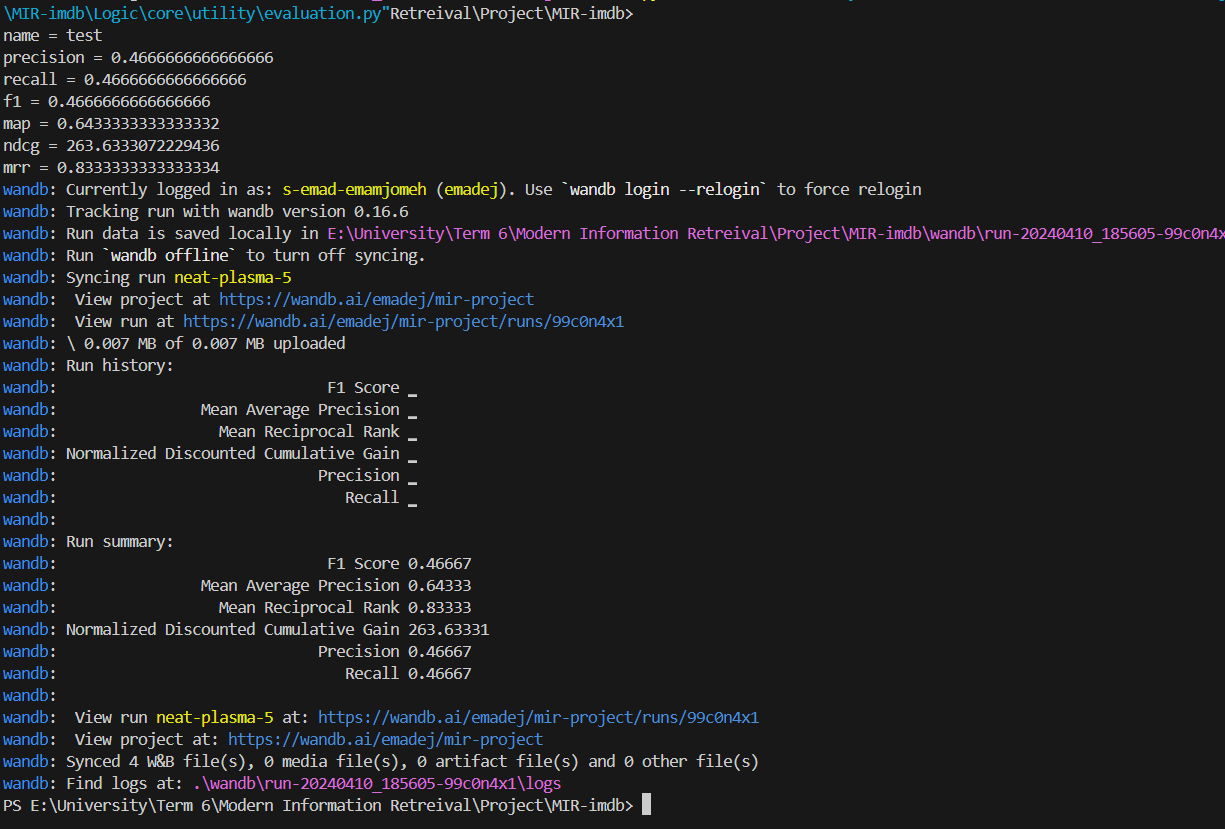
Here’s an example of spell corrections functionality:



# Evaluation

We run evaluation for 3 queries: (dune, spider-man spider-verse, matrix)

The metrics are as described below:



Wandb Page:

