623.610 – Information Search and Recommendation Systems	23S
Evaluation report	
Zentilin Xavier - Persico Loris	12040047 - 12244072

1 Overview

The goal of this project is to work with different strategies to generate recommendations of similar items, in this case movies. The first task is to develop a number of at least 5 functions in Python that, given a reference movie ID, return a ranked list of the top-5 most similar items. Each function has to implement a different strategy.

Next, implement a user interface, where the reference movie can be searched by title. The user can then select one of the returned search results and the system then presents five lists, each one containing similar movie recommendations. Each list is based on one of the implemented similarity functions. The final task lies in the evaluation of the different strategies

2 Recommendation Strategies

The five strategies we decided to use through these characteristics to recommend the different films: similarity of genres, similarity of keywords, similarity of actors, similarity of plot, similarity of synopsis.

Furthermore, for each list, after calculating the similarity and obtaining the list of recommendations, we filtered them using a popularity index to obtain the 5 least watched films, in order to try to provide recommendations from the long tail.

In particular for genre similarity, keyword similarity and actor similarity we used Kulczynski similarity, while for plot and synopsis strategies we transform texts into TF-IDF vectors and then calculate cosine similarity between They.

3 Evaluation

3.1 Study design

We conducted an online study using Google Forms on some users in which participants rated the quality of possible recommendations of 5 movie lists generated by different algorithms.

Participants received 5 films (Toy Story, The Godfather, The Terminator, The exorcist, Forrest Gump) and for each film they were presented with 5 different lists of films recommended by the different algorithms and were required to rank the different lists according to their preferences on which they would like more to receive in a real streaming site (ex. Netflix).

For each film presented, students are therefore asked to:

- Indicate if you know and/or have seen this film.

- Indicate the order of approval of the proposed lists.

To indicate the liking factor, the strategy we have recommended to users is to think personally about what types of recommendations you would most like to find in a movie streaming distribution system via the Internet (Netflix, for example).

Furthermore, each film was presented with a title, year of release and cover and for each film in the lists there was a link to the relative page of the "IMDb" site where all the details of the film are presented, in order to allow users a higher quality assessment about unknown movies.

3.2 Study outcomes

In total we reached 20 participants (all from Italy around the age of 20-30) with a percentage of total evaluations, given that there was the option not to evaluate films that were not known to each other, of 81

In our study we used two different rank aggregations: the Winning Frequency (WF) defined as how often the recommendations of a technique were considered the best by the user and the Borda Count (BC) used to aggregate the ranking of all five alternatives.

Strategy	Toy Story	Forrest Gump	The Godfather	The Exorcist	Terminator
A Genre Similarity	31,58%	27,78%	31,25%	23,08%	33,33%
B Keywords Similarity	52,63%	22,22%	25,00%	7,69%	33,33%
C Actors Similarity	5,26%	22,22%	12,50%	0,00%	0,00%
D Plot similarity	10,53%	5,56%	6,25%	15,38%	6,67%
E Synopsis similarity	0,00%	22,22%	18,75%	53,85%	26,67%

Winning Frequency

Rank	List	Formula	Points
1	A	N - 1	4
2	B	N - 2	3
3	C	N - 3	2
4	D	N - 4	1
5	E	N - 5	0

Borda Count method used

Strategy	Toy Story	Forrest Gump	The Godfather	The Exorcist	Terminator
A Genre Similarity	53	37	33	30	40
B Keywords Similarity	59	47	38	29	40
C Actors Similarity	38	46	27	21	18
D Plot similarity	20	29	33	14	24
E Synopsis similarity	20	21	19	36	28

Borda Count

Strategy	WF	BC
A Genre Similarity	**29,63%**	193
B Keywords Similarity	**29,63%**	**213**
C Actors Similarity	8,64%	150
D Plot similarity	8,64%	120
E Synopsis similarity	22,22%	124

Evaluation summary

Observing the results of the table on the winning frequency we can notice some interesting aspects, starting for example from strategy E where we notice a great appreciation for the Exorcist list which is made up of 2 sequels of the film, therefore we can deduce that users appreciate a lot when after having finished a film, one of the same saga is recommended.

Also for strategy C, as we can see in the tables, for 2 out of 5 films we got 0% winning frequency, we can explain this by looking at the recommended films in this list, they may have the same actor in common but it remains only a detail that will be considered relevant only to fans of that actor, while it can probably be considered irrelevant by most users since one actor can play so many different roles in even completely different movies.

A different but actually similar argument can be made for strategy D, we can explain this results by looking at the recommended films in this list, which may have something particular in common but by not taking into account the genre of the film the results are very approximate and almost always very far from the type of the initial film.

4 User Feedbacks

At the end of the evaluation form it was possible to leave written feedback indicating various observations, or what the evaluation criteria was for expressing one's preferences.

One recurring feedback was regarding the difficulty of making actual evaluations given the lack of knowledge of the proposed titles (which, however, was partially made up for by providing a link to a description page of each proposed film).

Other minor feedback concerned the judging criteria used, in particular one person gave more consideration to lists where there were films he had already seen and another relied on the year of the film's release.

From the feedback received we can think that recommending not popular items can be a good option to discover new things but it should be mixed with some known items for helping the user.

5 Conclusions

5.1 Which recommendation method led to the best results?

As we can see from the summary table, it is clearly seen that the two methods considered best (genre and keywords) are those based on generic labels that try to summarize their content. In particular, the method using keywords was slightly preferred, which we could describe as more detailed genres, confirming the effectiveness of this method.

5.2   Which method led to surprises?

The similarity of the synopsis was unexpected for us, as we did not think that recommending films with this technique could produce significant results since for each film a too specific characteristic is taken into consideration and not a generic description as can be represented by the genre, but still managed to win a good general appreciation.

5.3   Which method led to poor or unexpected result?

Basing recommendations on synopsis or plot can generate really bizarre results, such as the recommendation for Toy Story from Child's Play 2, it is true that they share toys but we do not think it is a good recommendation. In general, as previously discussed, the two worst methods (similarity of actors and plot) have often generated bizarre results given that, for example, an actor can star in films that are very different from each other and basing recommendations solely on this detail we did not think be (and has been confirmed) the winning choice. It could perhaps work in conjunction with one of the two best methods, i.e. also taking into consideration the type of film.