ANALYSIS

Part 1: Semantic Search Model

What type of queries tend to do well? Which is not so well?

The query 1 (Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions) and query 2 (Western Romance) tend to do well with 3 relevant and 2 irrelevant suggestions respectively.
The query 3 (Silent film about a Parisian star moving to Egypt, leaving her husband for a baron, and later reconciling after finding her family in poverty in Cairo) tends to do poorly with only 1 relevant and 4 irrelevant suggestions.
The query 4 (Comedy film, office disguises, boss's daughter, elopement), query 5 (Lost film, Cleopatra charms Caesar, plots world rule, treasures from mummy, revels with Antony, tragic end with serpent in Alexandria), query 6 (Denis Gage Deane-Tanner) performs mediumly with 2 relevant and 3 irrelevant suggestions respectively.

In a nutshell, the model performs well with generalized queries and does not tend to do well with precise queries.

For the queries that the model didn't perform well, what could be two alternative approaches?

The model did not perform well for queries 3 and 4 giving only one relevant and four irrelevant search results. It is because of giving all the precise details in the queries. For example, let us take a sample query "Love story, good ending takes place in France." If you give this kind of query, the search result obtained will be movies related to France, with good endings and love story films but the movie plot with all three elements is only relevant. So the results gained including only France or love stories or bad endings are irrelevant.
The above problem arises after using the "all-MiniLM-L6-v2" model approach. To solve the problem, we could use alternate approaches like

Hybrid Query Expansion with Controlled Filtering: This method involves expanding the initial user query to include synonyms or related terms and then applying controlled filtering to ensure the results meet all specified criteria. By expanding the query we can even get alternate words with the same meanings and with the controlled search, we can apply the filters strictly to get only relevant results.

Re-ranking with Cross Encoder: Along with the bi-encoder, there should be Re-ranking with Cross Encoder. They can perform a more detailed, interaction-based comparison between the query and each entry in the dataset, to get better relevance.
This method is effective for queries with multiple specific criteria because the cross-encoder can evaluate the composite meaning of the query against each document.

One thing to note is that it depends on the dataset too. With a dataset, the methods might vary too. If you have only one relevant item and instruct the model to bring five results then absolutely, it fetches the remaining four irrelevant search results as it intended to get five.

EXECUTION SUMMARY

Recommendation:

1. Semantic Search:

• Pros:
- Faster to execute and very simple and easy to implement.
- Can be straightforward to implement and scale for large datasets.

• Cons:
- Accuracy might not be that much greater when compared with RAG and Re-Ranked combining methods. It relies heavily on the quality of embeddings.
- It struggles while handling the data with multiple constraints (precise queries) like "Thriller, Germany, Dark".

2. Re-Ranked combining:

• Pros:
- It combines positives of BM25, Semantic Search along with combining ranked concept.
- It can even handle multiple constraints in the queries when used with Cross Encoders.

• Cons:
- Requires additional computational resources and complexity to integrate and manage multiple models. It takes lot of time
- As it combines the results, the effectiveness depends on the retrieval methods used initially.

3. RAG (Retrieval Augmented Generation):

• Pros:
- Has high accuracy due its core concept of sending a relevant document with a query to get better results as the model understands context too.
- Can produce more informative and contextually relevant responses compared to traditional generation models.

• Cons:
- Requires more complex architecture and high processing power. It takes a lot of time.
- It is very complex and difficult to implement especially when applying the concept to a manually uploaded dataset rather than the online datasets.

We would be recommending the method “Re- Ranked Methods” as Semantic Search was a simpler and very basic model but with a good amount of accuracy. But we believe the real world models must be a bit advanced to effectively retrieve and search as the datasets in the real world are very complex. Out of RAG and Re Ranked methods, implementing RAG was so difficult when dealing with the manually loaded dataset but it will not be a very big problem as many datasets will be available online. But when the server fails, it would create a huge mess. Re-Ranked concept covers BM25, Semantic Search. It will have both the advantages and even if it requires high processing powers and high end devices but many of the organizations can afford it as it does not require more processing power than RAG. It is a worthy risk to take.

Production: The model is successful as we are able to get the results with ease. But to deploy it into the real world it requires some additional things like choosing a cloud infrastructure to deploy like AWS, Google Cloud or Azure, any model needs further updates and concept of reusability. We did not upload everything according to structure. Everything must be clear so that everyone will be able to get it and try replicating it.

For these challenges might be scalability- due to large datasets, there might be a problem of scalability and we are planning to balance loads by implementing it across different servers. During handling the large datasets, it might be tough and expensive to but monitoring resources will save some amount. Still, we recommend going with the “Re-Ranked” model as it will incur profits in the long run and still be more cost effective than RAG.

RAGs/ LLMs: //how can you evaluate your model performance //how it can be compared with BM25, Semantic search, re ranked //factors affecting models performance
The results obtained from the model helps us to understand its performance. Even though it was complex, it gave relevant results. Almost for every result, it gave three relevant results except for the second and fifth query. Both were very precise queries with unique storyline. It gave at least one relevant result but BM25 gave all irrelevant results for one case.

Model’s performance can be evaluated by various metrics like Precision@k, Recall@k, MMR, F1 Score. In this project, we did only Recall@K and MMR. When we analyzed these models using the metrics and relevant search results. It was better than BM 25 and Semantic Search. It gave a neck fight with re-ranking but implementing this RAG with the manual dataset from a local device was difficult and the RAG theoretical concept was very simple. Query ambiguity and lack of five relevant search results to bring five might affect the model’s performance. Along with that, selecting appropriate methods like Mistral and datasets from the internet directly could have shown some different results.

Fine Tuning: There are many pre-trained models available for using in sentence transformers libraries like BERT, etc. The thing is that each model is trained on a particular dataset for accomplishing a particular task. When we pick one pre-trained model, we will lose a great amount of accuracy if that particular model is not trained on the dataset before. That's where Fine Tuning comes in. Fine tuning means training that particular model on the dataset we are going to use so that it gains some knowledge about the dataset and then as it already trained on that domain once, it can give better accuracies while we are doing the task. By this way, we can use any model of better accuracy to our project as we fine tune it before using it in the project so that it will have some knowledge about our domain prior to executing in the project.

APPENDIX

Part 1: Semantic Search Model

Query 1: “Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions”

1. Nanook of the North
2. The Frozen North
3. In the Land of the Head Hunters
4. Masked Emotions
5. Chang: A Drama of the Wilderness

Query 2: “Western romance”

1. Romance
2. Bucking Broadway
3. Wild and Woolly
4. A Romance of Happy Valley
5. The Enchanted Cottage

Query 3: "Silent film about a Parisian star moving to Egypt, leaving her husband for a baron, and later reconciling after finding her family in poverty in Cairo."

1. Sahara
2. Married in Hollywood
3. Mothers Cry
4. The House with Closed Shutters
5. A Busy Day

Query 4: "Comedy film, office disguises, boss's daughter, elopement."

1. Ask Father
2. Caught in a Cabaret
3. Pay Day
4. A Busy Day
5. Amarilly of Clothes Line Alley

Query 5: "Lost film, Cleopatra charms Caesar, plots world rule, treasures from mummy, revels with Antony, tragic end with serpent in Alexandria."

1. Cleopatra
2. 4 Devils
3. Disraeli
4. Bound in Morocco
5. Souls for Sale

Query 6: "Denis Gage Deane-Tanner"

1. Captain Alvarez
2. Hold Everything
3. Rookies
4. Near the Rainbow’s End

5. A man from Wyoming

Part 2: Reranker notebook

1. **BM25 Method**

Query 1: “Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions”

1. I Do
2. Tarzan of the Apes
3. The Scar of Shame
4. Kiki
5. Hell Harbor

Query 2: “Western romance”

1. The Call of the Wild
2. Wild and Woolly
3. Romance
4. Four Sons
5. The Forbidden City

Query 3: "Silent film about a Parisian star moving to Egypt, leaving her husband for a baron, and later reconciling after finding her family in poverty in Cairo."

1. Sahara
2. He Who Gets Slapped
3. Broken Hearts of Hollywood
4. Peacock Alley
5. True Heart Susie

Query 4: "Comedy film, office disguises, boss's daughter, elopement."

1. The Boy Friend
2. Mable’s Blunder
3. Bucking Broadway
4. Cruel, Cruel Love
5. A Busy Day

Query 5: "Lost film, Cleopatra charms Caesar, plots world rule, treasures from mummy, revels with Antony, tragic end with serpent in Alexandria."

1. Cleopatra
2. Mama’s Affair
3. Peter Pan
4. Madame X
5. This Exquisite Thief

Query 6: "Denis Gage Deane-Tanner"

1. Captain Alveriz
2. One Week

3. Old Lady 31
4. Number, Please ?
5. Now or Never

2. BM-25 and Retrieval, Re-ranked both combined

Query 1: “Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions”

1. Mamba
2. The Four Horsemen of the Apocalypse
3. Tarzan of the Apes
4. Atlantis
5. Chang: A Drama of the Wilderness

Query 2: “Western romance”

1. The General
2. Romance
3. The Sheikh
4. Wild and Woolly
5. All Quiet on the Western Front

Query 3: "Silent film about a Parisian star moving to Egypt, leaving her husband for a baron, and later reconciling after finding her family in poverty in Cairo."

1. Sahara
2. A Women of the Affairs
3. He Who Gets Slapped
4. Foolish Wives
5. Kiki

Query 4: "Comedy film, office disguises, boss's daughter, elopement."

1. Ask Father
2. Bucking Broadway
3. Mable’s Blunder
4. His Wedding Night
5. Showboat

Query 5: "Lost film, Cleopatra charms Caesar, plots world rule, treasures from mummy, revels with Antony, tragic end with serpent in Alexandria."

1. Cleopatra
2. Mama’s Affair
3. The Hunchback of Notre Dame
4. Three Ages
5. What Daisy Said

Query 6: "Denis Gage Deane-Tanner"

1. Captain Alvarez

2. Silk Husbands and Calico Wives
3. Hangman’s House
4. The Law of Men
5. The Delicious Little Devil

Part 3: Retrieval Augmented Generation (RAG)

Query 1: “Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions”

1. Nanook of the North
2. In the Land of the Head Hunters
3. The Frozen North
4. From Leadville to Aspen
5. Chang

Query 2: “Western romance”

1.Romance
2.Wild and Woolly
3.Bucking Broadway
4.The Enchanted Cottage
5.A Romance of Happy Valley

Query 3: "Silent film about a Parisian star moving to Egypt, leaving her husband for a baron, and later reconciling after finding her family in poverty in Cairo."

1.Sahara
2.Mothers Cry
3.The House with Closed Shutters
4.A Busy Day
5.The Suburbanite

Query 4: "Comedy film, office disguises, boss's daughter, elopement."

1.Ask Father
2.Caught in a Cabaret
3.The Extra Girl
4.Mabel's Blunder
5.Amarilly of Clothes-Line

Query 5: "Lost film, Cleopatra charms Caesar, plots world rule, treasures from mummy, revels with Antony, tragic end with serpent in Alexandria."

1.Cleopatra
2.A Daughter of the Gods
3.Disraeli
4.A Splendid Hazard
5.The Sorrows of Satan

Query 6: "Denis Gage Deane-Tanner"

1. Captain Alvarez
2. Near the Rainbow's
3. A Man from Wyoming
4. The Wolf Song
5. Tenderloin

Part 4: All in one Data Table

Query 1	Recall@1	Mean Reciprocal Rank
Semantic Search	0.33	1
BM25	0	0
Reranker	0.33	1
RAG	0.33	1

Query 2	Recall@1	Mean Reciprocal Rank
Semantic Search	0.33	1
BM25	0.25	1
Reranker	0.33	1
RAG	0.33	1

Query 3	Recall@1	Mean Reciprocal Rank
Semantic Search	1	1
BM25	0.5	1
Reranker	0.5	1
RAG	0.5	1

Query 4	Recall@1	Mean Reciprocal Rank
Semantic Search	0.5	1
BM25	0	2
Reranker	0.5	1
RAG	0.5	1

Query 5	Recall@1	Mean Reciprocal Rank
Semantic Search	0.5	1
BM25	1	1
Reranker	1	1
RAG	1	1

Query 6	Recall@1	Mean Reciprocal Rank
Semantic Search	0.5	1
BM25	1	1
Reranker	0.5	1
RAG	0.5	1