

Homework 3-Manjinder Sandhu

Analysis Writeup

Task: The goal of this homework assignment is to create a prototype (POC/MVP) search tool that will help users find old 1920s movies. The movie dataset will contain 1,000 movies.

The three models that I used were Semantic Search, Retrieve & Re-Rank, and Retrieval Augmented Generation (RAG). Semantic Search uses a natural language process that understands the meaning of your search which can provide better accuracy. In Nils Reimers article he states that “Semantic search seeks to improve search accuracy by understanding the content of the search query. In contrast to traditional search engines which only find documents based on lexical matches”. There are many pros and cons to using a Semantic Search. The biggest pro is that semantic search uses synonym recognition which can enhance the search experience. A con to semantic search is that it can be very time-consuming to get a search result. Retrieve & Re-Rank is where the initial search results are checked again by a cross-encoder which gives a score. The BM25 will pick out movies based on related keywords while the rerank will ensure it is the best overall fit. The higher the scores the better the results. The biggest pro for using a retrieve & re-rank model is that a cross-encoder can improve your search results. The article, which will be linked in the Appendix states “A re-ranker based on a Cross-Encoder can substantially improve the final results for the user”. A con for the retrieve & re-rank model is that the result might not be relevant to the search. RAG is where we use large language models to get accurate results. The pro of RAG is that it RAG is very easy to implement. In this homework, we just had to get a firework API for RAG LLM to work. The con for RAG is that it requires reliance on external sources. In the homework, we relied on Firework API for our LLM to generate results. I recommend we use the retrieve & re-rank model because they provide the best Recall@1 and Mean Reciprocal Rank for BM25 and rerank compared to Semantic Search. BM25 had the highest score. We wouldn’t have to rely on APIs and other external sources to run RAG LLM. BM25 can quickly find relative research while the rerank can ensure the best overall fit.

To deploy the Retrieve & Re-Rank model we first have to ensure that use the right GPU. In the assignment, we only used 1,000 movie titles and plots. If we use a large complex dataset, it may require us to purchase Colab Pro+, Pay As You Go plan, or Colab Enterprise because the

free tire may not support that. Another thing we need for the model to run is we will need to download the different packages like sentence transformation. The packages have many pre-trained models that we will use but that requires us to verify that we are using the correct pre-trained model. The biggest challenge that we may face is writing the Python code correctly for the model to run and produce results. I used ChatGPT to help me write and explain the Python code for all three models. ChatGPT was not always correct and it did require me to go back and correct its mistakes in the code. It is a requirement to have human intervention when using ChatGPT. I still recommend that we stick to the Retrieve & Re-Rank model and it doesn't alter my recommendations. The model provided very good results when we used it on the 1,000 movie dataset.

When I used RAG LLM my results did not have any of the ground truth movies in the search results. The model could not match the correct movie when it was given a query. I believe the reason why this happened was because the model may have had a misunderstanding or incorrect interpretation due to the prompt. Recall@1 and Mean Reciprocal Rank can be used to evaluate Semantic Search, BM25, and Rerank. It can't be used for RAG LLM because the model generates text. For RAG LLM we may have to use manual evaluation like looking at the results and see if they contain the ground truth. A major factor that can affect the model performance is the quality of the external source. We relied on Firework API for our LLM to generate results but if the Firework didn't have high-quality and up-to-date data then it could affect our results.

Fine-tuning is a process where we have a pre-trained model and we train the pre-trained model with examples that are relevant to the task that we want the model to perform on. For us to approach this problem we would first need to choose a pre-trained model that is good with movie data. Then we would want to train the pre-trained model with specific movie examples. We want the model to find patterns. Then we can evaluate the fine-tuned model with recall to measure how well it performed. Based on the evaluation we can refine the model to improve its performance. In the article, it says "Fine-tuning techniques help optimize the model's performance for targeted applications". I believe that it is worth the effort to fine-tune a pre-trained model and I would recommend it to my boss.

In semantic search, queries with similar lengths and content perform well like symmetric semantic search. Asymmetric semantic search are questions like queries for example we may ask "Can you summarize the movie?". In our homework, a symmetric semantic search may perform

well because we do not desire long paragraph-type results. If we have queries that the model didn't perform well on we can fine-tune the model to improve its performance or we can also switch our model and implement a retrieve & rerank pipeline instead.

Appendix

Semantic Search Mode:

Query	Ground Truth Movie	Your Results
"Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions"	'Nanook of the North'	<ol style="list-style-type: none"> 1. Nanook of the North 2. David Copperfield 3. Straight Shooting 4. The Frozen North 5. The Boy Friend
"Western romance"	'The Lucky Horseshoe'	<ol style="list-style-type: none"> 1. The Great Gatsby 2. Youth's Endearing Charm 3. Frankenstein 4. High Society Blues 5. Romance
"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."	"Sahara"	<ol style="list-style-type: none"> 1. Married in Hollywood 2. The King on Main Street 3. Sahara 4. One Hysterical Night 5. Bound in Morocco
"Comedy film, office disguises, boss's daughter, elopement."	"Ask Father"	<ol style="list-style-type: none"> 1. Youth's Endearing Charm 2. Dressed to Kill 3. The Boy Friend 4. A Little Journey 5. This is Heaven
"Lost film, Cleopatra charms Caesar, plots world rule, treasures from mummy, revels with Antony, tragic end with serpent in Alexandria."	"Cleopatra"	<ol style="list-style-type: none"> 1. Cleopatra 2. The Lost World 3. The Man Who Lost Himself 4. The Golden Louis 5. Reaching for the Moon
"Denis Gage Deane-Tanner"	"Captain Alvarez"	<ol style="list-style-type: none"> 1. Souls for sale 2. The Man in Possession 3. The Man from Blankley's 4. Blind Youth 5. Old Clothes

BM25:

Query	Ground Truth Movie	Your Results
"Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions"	'Nanook of the North'	<ol style="list-style-type: none">1. I Do2. Uncharted Seas3. Powers That Prey4. Rough Waters5. The King of Kings
"Western romance"	'The Lucky Horseshoe'	<ol style="list-style-type: none">1. The Call of the Wild2. Romance3. Four Sons4. A Man's Fight5. Gun Smoke
"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."	"Sahara"	<ol style="list-style-type: none">1. Sahara2. He Who Gets Slapped3. Inspiration4. A Lad from Old Ireland5. The Maltese Falcon
"Comedy film, office disguises, boss's daughter, elopement."	"Ask Father"	<ol style="list-style-type: none">1. Mabel's Blunder2. Bucking Broadway3. Ask Father4. The Front Page5. His Wedding Night
"Lost film, Cleopatra charms Caesar, plots world rule, treasures from mummy, revels with Antony, tragic end with serpent in Alexandria."	"Cleopatra"	<ol style="list-style-type: none">1. Cleopatra2. Captain Applejack3. Fair Lady4. Reaching for the Moon5. Tom Sawyer
"Denis Gage Deane-Tanner"	"Captain Alvarez"	<ol style="list-style-type: none">1. Captain Alvarez2. Hangman's House3. Through the Back Door4. SureFire5. The Sky Pilot

Re-Rank Combined:

Query	Ground Truth Movie	Your Results
"Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions"	'Nanook of the North'	<ol style="list-style-type: none"> 1. David Copperfield 2. Nanook of the North 3. The Frozen North 4. Straight Shooting 5. The Salvation Hunters
"Western romance"	'The Lucky Horseshoe'	<ol style="list-style-type: none"> 1. Romance 2. Married 3. The Great Gatsby 4. Frankenstein 5. Youth's Endearing Charm
"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."	"Sahara"	<ol style="list-style-type: none"> 1. Sahara 2. Morocco 3. Married in Hollywood 4. The Golden Louis 5. The King on Main Street
"Comedy film, office disguises, boss's daughter, elopement."	"Ask Father"	<ol style="list-style-type: none"> 1. Bumping into Broadway 2. The Pasha's Daughter 3. Youth's Endearing Charm 4. A Little Journey 5. Dressed to Kill
"Lost film, Cleopatra charms Caesar, plots world rule, treasures from mummy, revels with Antony, tragic end with serpent in Alexandria."	"Cleopatra"	<ol style="list-style-type: none"> 1. Cleopatra 2. The Lost World 3. The Man Who Lost Himself 4. Reaching for the Moon 5. The Golden Louis
"Denis Gage Deane-Tanner"	"Captain Alvarez"	<ol style="list-style-type: none"> 1. The Man from Blankley's 2. Blind Youth 3. The Blot 4. Old Clothes 5. Caught Plastered

RAG:

Query	Ground Truth Movie	Your Results
"Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions"	'Nanook of the North'	<ol style="list-style-type: none">1. Arctic Survival: A Journey Through Indigenous Communities2. Life in the Frost: The Arctic's Indigenous Peoples3. Beyond the Ice: The Lives of Arctic Indigenous Communities4. Arctic Resilience: The Survival of Indigenous Peoples5. Arctic Voices: The Stories of Indigenous Peoples in the North
"Western romance"	'The Lucky Horseshoe'	<ol style="list-style-type: none">1. The Last Cowboy2. Rustling Romance3. The Wild West Love Story4. A Gunfighter's Heart5. The Long Ranger's Lady
"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."	"Sahara"	<ol style="list-style-type: none">1. Egyptian Dream2. Parisian Passion3. Cairo's Call4. Love in the Desert5. The Return to Paris
"Comedy film, office disguises, boss's daughter, elopement."	"Ask Father"	<ol style="list-style-type: none">1. Office Escapade: A Comedy of Disguises2. The Boss's Daughter's Elopement: A Comedy3. The Office Elopement: A Comedy of Errors4. The Comedy of Disguises: An Office

		Elopement 5. The Elopement of the Boss's Daughter: A Comedy
"Lost film, Cleopatra charms Caesar, plots world rule, treasures from mummy, revels with Antony, tragic end with serpent in Alexandria."	"Cleopatra"	1. The Cleopatra Conspiracy 2. Caesar's Charm 3. The Rise of Cleopatra 4. The Mummy's Treasure 5. Antony's Reveal
"Denis Gage Deane-Tanner"	"Captain Alvarez"	1. The Deane-Tanner Chronicles 2. Deane-Tanner: A Family Legacy 3. Deane-Tanner: The Next Generation 4. Deane-Tanner: A Legacy Continues 5. Deane-Tanner: The Legacy Lives On

Experiment Results:

	Recall@1	Mean Reciprocal Rank
Semantic Search	0.333	0.3833
BM25	0.5	0.55
Rerank	0.33	0.416

Citations:

<https://www.sbert.net/examples/applications/semantic-search/README.html>

https://www.sbert.net/examples/applications/retrieve_rerank/README.html#retrieve-re-rank