

## HW3 Write up / Appendix

### Recommendation:

Out of the 3 models used for this experiment, I would have to say that the Reranker model was the one that performed the best. Semantic search was the simplest, however it was less accurate overall in returning the ground truth movie. BM25 gave better results for queries that had slightly more detailed descriptions but was not as accurate with shorter broader ones. Reranker on the other hand returned the ground truth movie on 5 out of 6 examples, with the only one being the short “western romance” description.

### Production:

The first consideration we could think about is how structured our data is. Are there any duplicates or things that could potentially throw an error or give the user the wrong information? The next thing would be to know the context of how the data is going to be used. For example, a search query for a movie and its plot can be very different than a query that is looking for a website. Considering we have a use case where we need keyword matches, The reranker would still be my recommendation based on its ability to understand the context of the words being used over the lexical and semantic searches. There is one drawback with using the reranker, and that is that it is much more resource demanding than the other two models.

### RAG/LLM's:

The RAG model did not perform as well as any of the other three models. It often did not return the correct movie and would return duplicates as well. Some of the factors that lead to this was that it involves a prompt that instructs the model on what and how to return. Very small changes to the prompt could lead to very different results, so figuring out what the most effective instruction is likely to be difficult.

### Fine Tuning:

Assuming we are not using pretrained models, we would have to develop our own model that is trained to best suite our needs. This could take some time and money to develop depending on how detailed and complex we want the system to be. Once we have fit and adjusted our model, we just have to worry about optimizing it further.

## Results:

	Recall @ 1	Mean Reciprocal Rank
Semantic Search	0.33	0.38
BM25	0.5	0.55
Reranker	0.67	0.67

## Semantic Search Output:

"Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions"	'Nanook of the North'	<ol style="list-style-type: none"> <li>1. 0.457 - Nanook of the North</li> <li>2. 0.264 - David Copperfield</li> <li>3. 0.251 - Straight Shooting</li> <li>4. 0.244 - The Frozen North</li> <li>5. 0.237 - The Boy Friend</li> </ol>
"Western romance"	'The Lucky Horseshoe'	<ol style="list-style-type: none"> <li>1. 0.305 - The Great Gatsby</li> <li>2. 0.295 - Youth's Endearing Charm</li> <li>3. 0.283 - Frankenstein</li> <li>4. 0.280 - High Society Blues</li> <li>5. 0.277 - Romance</li> </ol>
"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"> <li>1. 0.464 - Married in Hollywood</li> <li>2. 0.433 - Sahara</li> <li>3. 0.416 - The King on Main Street</li> <li>4. 0.411 - Peacock Alley</li> <li>5. 0.409 - Bound in Morocco</li> </ol>
"Comedy film, office disguises, boss's daughter, elopement."	'Ask Father'	<ol style="list-style-type: none"> <li>1. 0.389 - Youth's Endearing Charm</li> <li>2. 0.362 - The Boy Friend</li> <li>3. 0.350 - The Pasha's Daughter</li> <li>4. 0.350 - The Evidence of the Film</li> <li>5. 0.340 - Mabel's Blunder</li> </ol>
"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"> <li>1. 0.537 - Nanook of the North</li> <li>2. 0.365 - The Lost World</li> <li>3. 0.358 - The Man Who Lost Himself</li> <li>4. 0.353 - The Golden Louis</li> <li>5. 0.337 - Reaching for the Moon</li> </ol>
"Denis Gage Deane-Tanner"	'Captain Alvarez'	<ol style="list-style-type: none"> <li>1. 0.365 - Souls for Sale</li> <li>2. 0.363 - The Man in Possession</li> <li>3. 0.344 - The Man from Blankley's</li> <li>4. 0.333 - Blind Youth</li> <li>5. 0.303 - Old Clothes</li> </ol>

BM25 Output:

"Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions"	'Nanook of the North'	<ol style="list-style-type: none"> <li>1. 9.184 - I Do</li> <li>2. 8.223 - Uncharted Seas</li> <li>3. 6.452 - Powers That Prey</li> <li>4. 4.968 - Rough Waters</li> <li>5. 4.424 - The King of Kings</li> </ol>
"Western romance"	'The Lucky Horseshoe'	<ol style="list-style-type: none"> <li>1. 6.360 - The Call of the Wild</li> <li>2. 5.661 - Four Sons</li> <li>3. 5.591 - Romance</li> <li>4. 5.504 - A Man's Fight</li> <li>5. 5.428 - Gun Smoke</li> </ol>
"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"> <li>1. 30.465 - Sahara</li> <li>2. 16.394 - He Who Gets Slapped</li> <li>3. 14.060 - Inspiration</li> <li>4. 13.931 - A Lad from Old Ireland</li> <li>5. 12.697 - The Maltese Falcon</li> </ol>
"Comedy film, office disguises, boss's daughter, elopement."	'Ask Father'	<ol style="list-style-type: none"> <li>1. 13.147 - Mabel's Blunder</li> <li>2. 9.583 - Bucking Broadway</li> <li>3. 8.566 - Ask Father</li> <li>4. 8.252 - The Front Page</li> <li>5. 7.900 - His Wedding Night</li> </ol>
"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"> <li>1. 70.468 - Cleopatra</li> <li>2. 10.867 - Captain Applejack</li> <li>3. 9.121 - Fair Lady</li> <li>4. 8.343 - Reaching for the Moon</li> <li>5. 8.258 - Tom Sawyer</li> </ol>
"Denis Gage Deane-Tanner"	'Captain Alvarez'	<ol style="list-style-type: none"> <li>1. 28.985 - Captain Alvarez</li> <li>2. 4.114 - Hangman's House</li> <li>3. 0.000 - Red Courage</li> <li>4. 0.000 - The Sea Lion</li> <li>5. 0.000 - A Sailor-Made Man</li> </ol>

Reranker Output:

"Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions"	'Nanook of the North'	<ol style="list-style-type: none"> <li>1. -4.554 – Nanook of the North</li> <li>2. -10.989 – Ravished Armenia</li> <li>3. -11.129 - Experience</li> <li>4. -11.183 – From Leadville to Aspen</li> <li>5. -11.191 – Masked Emotions</li> </ol>
"Western romance"	'The Lucky Horseshoe'	<ol style="list-style-type: none"> <li>1. -4.111 – Romance</li> <li>2. -5.683 – Song of the West</li> <li>3. -5.739 – A Romance of Happy Valley</li> <li>4. -7.545 – Way Out West</li> <li>5. -9.987 – Love me and the World Is Mine</li> </ol>
"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"> <li>1. 4.676 - Sahara</li> <li>2. -5.127 - Cleopatra</li> <li>3. -5.283 – The Circle</li> <li>4. -8.483 – The Italian</li> <li>5. -8.686 – The Pasha's Daughter</li> </ol>
"Comedy film, office disguises, boss's daughter, elopement."	'Ask Father'	<ol style="list-style-type: none"> <li>1. 3.279 – Ask Father</li> <li>2. -0.478 – Mabel's Blunder</li> <li>3. -4.198 – His Wedding Night</li> <li>4. -5.658 – Amarilly of Clothes-Line Alley</li> <li>5. -6.670 - It</li> </ol>
"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"> <li>1. 6.336 -Cleopatra</li> <li>2. -6.723 – Youths Enduring Charm</li> <li>3. -9.085 – A Daughter of the Gods</li> <li>4. -9.763 – Adam and Evil</li> <li>5. -10.158 – The Pasha's Daughter</li> </ol>
"Denis Gage Deane-Tanner"	'Captain Alvarez'	<ol style="list-style-type: none"> <li>1. 6.481 - Captain Alvarez</li> <li>2. -10.600 – The Saphead</li> <li>3. -10.618 – Numbered Men</li> <li>4. -11.241 – Bucking Broadway</li> <li>5. -11.343- Branded</li> </ol>

RAG Output:

"Documentaries showcasing indigenous peoples' survival and daily life in Arctic regions"	'Nanook of the North'	<ol style="list-style-type: none"> <li>1. The Frozen North</li> <li>2. The Viking</li> <li>3. Call of the Wild</li> <li>4. The Ghost of Slumber Mountain</li> <li>5. From Leadville to Aspen</li> </ol>
"Western romance"	'The Lucky Horseshoe'	<ol style="list-style-type: none"> <li>1. Th Road to Romance</li> <li>2. Song of the West</li> <li>3. A Romance of Happy Valley</li> <li>4. Romance</li> <li>5. Romance</li> </ol>
"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"> <li>1. A Parisian Star in Egypt</li> <li>2. The Parisian Star</li> <li>3. The Parisian Star in Hollywood</li> <li>4. The Parisian Star in Egypt</li> <li>5. The Parisian Star in Cairo</li> </ol>
"Comedy film, office disguises, boss's daughter, elopement."	'Ask Father'	<ol style="list-style-type: none"> <li>1. Behind Office Doors</li> <li>2. A Rogue's Romance</li> <li>3. Beauty and the Rogue</li> <li>4. Time, the Comedian</li> <li>5. The Probation Wife</li> </ol>
"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"> <li>1. Cleopatra</li> <li>2. The Mummy</li> <li>3. Antony and Cleopatra</li> <li>4. The Lost City of Atlantis</li> <li>5. The Serpent Queen</li> </ol>
"Denis Gage Deane-Tanner"	'Captain Alvarez'	<ol style="list-style-type: none"> <li>1. Beau Geste</li> <li>2. Beau Brummel</li> <li>3. The Bondman</li> <li>4. Gentlemen of Nerve</li> <li>5. Beau Bandit</li> </ol>