

CS410 Final Project Progress Report

Hannah Benig - NETID: hhlim2

I have been having trouble getting data from the articles returned by NewsCatcherAPI. Sometimes the function I use to pull the text is blocked; other times the API itself was not returning articles from the queries I was entering. I have pivoted from using this API to instead querying my CS441 textbook.

1) Which tasks have been completed?

- Understand how the NewsCatcherAPI works and set it up to gather articles and metadata based on a query. This is no longer applicable to this project but I spent a long time trying to make it work. NewsCatcherAPI was not returning articles for all the queries I entered even when removing stopwords from the queries to only enter keywords.
 - I am now using my CS441 textbook as the corpus(10 hours)
- Set up LLM to receive context from textbook (1 hour)
- Evaluate LLMs from hugging face that we will subset with the news articles based on the answers to evaluation questions and ease of use (plan to try out Llama-2, Flan UL2, and roBERTa to see which works best) (5 hours)

2) Which tasks are pending?

- Create a set of evaluation questions (1 hour)
- Create vector database implementing BM25 for ranking (4 hours) - now only using BM25 ranking, see note in challenges below.
- Fine-tune LLM parameters (temperature, top-p, top-k, etc.) to find which combination gives the best answer (3 hours)
- Create Streamlit web app for frontend and package everything up for grading/evaluation (4 hours)

3) Are you facing any challenges?

- I ran into an issue with the Llama-2 and Flan UL2 models, they required more compute resources than my computer could handle. I have been able to get around this issue using a smaller model. I am working to provide the model with better information in order to return correct answers.
- BM25 ranking is not commonly paired with a vector database. Instead, I am using BM25 to rank passages from the textbook.