Alejandro Rocas
Lomis Chen
Bahaar Bhatia
Nikhilesh Sivapatham

Project Progress Report

Below is information on our team for the IR competition.

Progress Made

So far, our group has been able to parse through the json files and augment the data from the metadata csv file. We have placed the predicted relevance scores for the queries in the predictions text file. The text file submits the score for the top 1k documents per query and the ranker we used was okapibm25. We are in the process of improving the score. Our group has expanded the queries to capture the query, question, and narrative. Additionally, by adding more stopwords to the stopwords list, the highest score we were able to obtain so far is 0.6155538743352519.

Challenges & Remaining Tasks

The main challenge and remaining task includes improving the accuracy. Two attempts to improve accuracy that we implemented were (1) adding the intro and (2) using floats instead of integers data types to look at the score. When the intro was added and/ the data type was changed, the accuracy went down. Thus, we predicted that there may have been more noise being added with the intro. Some ways to increase accuracy that we plan to implement may be tokenizing.