

Alejandro Rocas  
Lomis Chen  
Bahaar Bhatia  
Nikhilesh Sivapatham

## IR Comp Project Progress Report

### Progress Made

Our group has overcome several key milestones since the beginning of the project. To start off, we first leveraged the code from MP2.2 as our basis – this includes replacing the content of qrels, queries, and metadata with the corresponding IR content. At the same time we also built off of the sample code from the original IR comp repo to i) convert the queries XML to txt format; and ii) extract the author, abstract, and relevant content from the metadata file. With those as the project foundation, we proceeded with computing the predictions, as required. The results were in the ~0.6s, which was slightly lagging the baseline and hence we decided to augment the data further by including the JSON content in hope to further improve the prediction scoring. We've been doing some tuning here and there throughout the process, and to date the highest score we were able to obtain is 0.6155538743352519.

For additional context, we used okapibm25 as our ranker. During the process we also attempted to augment the queries to include all *query*, *question* and *narrative* which slightly improved our prediction performance. Lastly, we experimented with including more stopwords to the stopwords list which delightfully gave a boost of +0.01 to our predictions

### Challenges & Remaining Tasks

The main challenge we are to overcome is pinpointing on the parts of the system we are to refine that'd help us improve the overall prediction accuracy. There are plenty approaches we could further experiment with. One would be augmenting the data with additional content, whether it be contents beyond the "intro" section or web-scraping additional content from other sites. Another approach is to tokenize the initial query, that may improve the scoring of relevance matching.

All in all, what remains for our group is to be even more creative and experiment with methods that'd bring us closer to the baseline