

MacEwan University at the TREC 2020 Fair Ranking Track

Brian Almquist (almquistb@macewan.ca)

MacEwan University

The MacEwan University School of Business submitted two runs for the TREC 2020 Fair Ranking Track. For this task, we indexed the document abstracts and the associated metadata from the provided Semantic Scholar dataset into a single Solr¹ node using a standard Tokenizer chain.

For each of the evaluation queries, we executed a query for each of the documents that required reranking. For each query-document pairing, we collected the BM25 similarity score for the “paperAbstract” and “title” fields. Each of the documents are ranked for each field based on the similarity score with the query text, with ties sharing their combined rank. This resulted in two ranked lists for each query.

For the baseline run, **MacEwan-base**, the reranking occurs through merging the two ranked lists using a weighted sum of the ranks.² For the normalized run, **MacEwan-norm**, the weighted sum uses the normalized similarity scores for each field instead of the similarity rank. Each sequence is assigned a set of starting weights. As queries are repeated within the sequence, the weights are adjusted incrementally to allow for emphasis or de-emphasis on different ranked lists. The amount of the adjustment is based on the frequency of the query within the sequence. This ensures that a complete shift in the weights for a given query will occur over the sequence if the query is repeated at least once. With only two ranked fields, the weights start equalized, and then the weight of the paperAbstract rank/score is increased while the weight of the title is decreased.

Due to time constraints, we were not able to incorporate information from the training grouping and any potential alternative groupings. As a result, these runs only assess whether fairness goals can be achieved with simple perturbations of result, given that the documents have met some threshold for relevance based on query similarity.

Run	Query Minimums	Query Maximums	Freq < Median	Mean over Topics
MacEwan-base	17	5	95	0.0349
MacEwan-norm	13	23	77	0.0993

We note that, while not universally inferior, the normalized run does not perform as well as the baseline run. Assuming a goal of minimizing the difference from the expected exposure, the baseline run achieves a better score in 54% of the queries. The difference in the mean scores between the two runs suggests that the distribution of similarity scores used for reranking the normalized run can lead to greater deviation

¹ <http://lucene.apache.org/solr/>

² Shaw, J. A. and Fox, E. A. Combination of multiple searches. In Harman, D. K., ed. *The Second Text REtrieval Conference (TREC-2)*. (Gaithersburg, MD, August 31-September 2, 1993). The Department of Commerce and the National Institute of Standards and Technology (NIST), 1993, 243-252.

from expected exposure. One possibility is that using a linear adjustment of weights on a linear distribution of scores, such as ranks, may produce a more frequent reranking than on normalized similarity scores. If there is a large gap in similarity scores, greater adjustments in weights will be required to force a reranking of the documents.

Again, other document metadata, such as citation numbers, or external information gathered about the authors, may also be used as inputs for these merged rank scores.