1. MAP / RPREC / NDCG@20 Results

The following are the metrics obtained (using trec_eval) for the page-level and section-level queries:

Method	MAP	RPREC	NDCG@20
page-bm25	0.6016	0.5966	0.7696
page-lnc.ltn	0.5222	0.5479	0.6828
page-bnn.bnn	0.5419	0.5358	0.7168
page-anc.apc	0.5591	0.5721	0.7268
section-bm25	0.3122	0.2347	0.4169
section-lnc.ltn	0.1889	0.1132	0.2887
section-bnn.bnn	0.2735	0.2055	0.3684
section-anc.apc	0.2430	0.1657	0.3455

Unsurprisingly, bm25 works the best in both instances. Also, the metrics for all methods drop down considerably when evaluated on section-level queries and hierarchical qrels. This is expected, as these kinds of queries are considered more "difficult".

Among the TFIDF variants we implemented: it appears that anc.apc did the best for page-level queries, but that simplistic bnn.bnn did best for the section-level queries! I am not sure why this is the case.

A couple of notes on our implementation:

- We were not able to come up with proper cosine normalization metric for the document vector. Instead we divided by the square of the document length when cosine was called for.
- We were not able to implemented the augmented term frequency for the document vector (although we were able to do this with the query vector). In light of this, it's surprising that anc.apc did so well when evaluated on the article queries.

2. Spearman Correlations

The following table describes the correlations of each of our TFIDF variants to that of the BM25 baseline. Methods prefixed with "page-" are obtained using article-level queries. Methods prefixed with "section-" are obtained using section-path queries (and using the hierarchical qrels).

Method	Correlation to BM25
page-lnc-ltn	0.7879634400797157
page-bnn-bnn	0.06518579123798646
page-anc-apc	0.5942227615001298
section-lnc-ltn	0.4568758152268715
section-bnn-bnn	-0.2519283633235704
section-anc-apc	0.38836993318712215

It appears that the lnc.ltn method is the most highly correlated with BM25 (when considering the article-level and section-level queries). Bnn does the worst, but then again that was a really stupid method so I'm not surprised... It's interesting, though, that it resulted in a negative correlation.

2.1. Resolving Missing Data

When a paragraph is absent in the BM25 query or one of the TFIDF variants, we compute its distance by taking the maximum distance among all other ranked paragraphs in the query, rather than omitting these paragraphs.

This could perhaps be the source of the negative correlation we are seeing in bnn.bnn, because if bnn.bnn contains fewer paragraphs in common the BM25, many of the distances would be calculated using the maximum distance method above. Why this is a problem is that it would imply a negative correlation when there may not actually be one!

3. Section Queries

As I have shown in sections 1 and 2, I also included the evaluations results and correlation results of section queries.

How I did it: I used the flatSectionPaths method on Data.Page to return the sections (list of lists). I concatenated the section headings of the lists, and the sections ids of the lists, to create the resulting section queries and their associated ids.

Note that, at least in the version of TREC Tools that we are using, the sections paths omit the article's page name and page id, so you need to add those to the beginning of each of the paths, and in addition also add one more query that is just the article-level query.