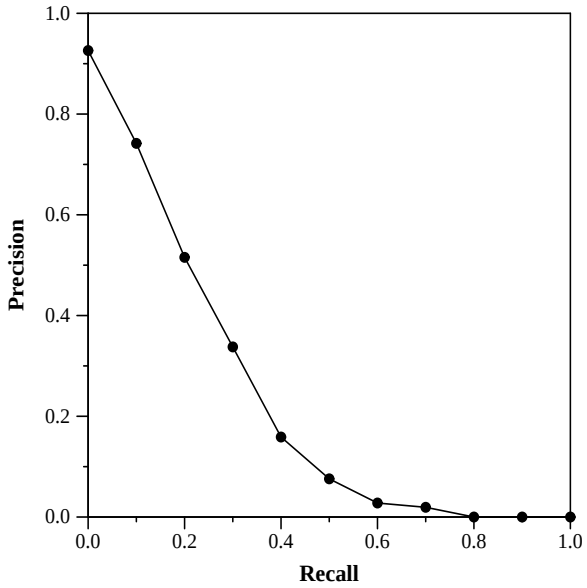
Deep Learning Track results — (ielab) The University of Queensland

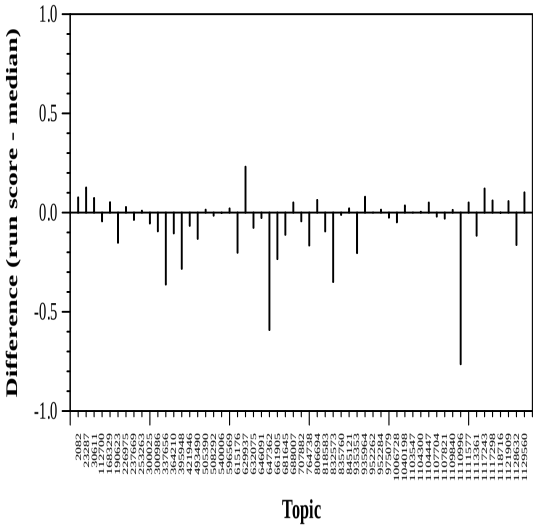
Run Description	
Run ID:	ielab-TILDEv2d
Task:	Document Ranking
Subtask:	Full ranking from the collection
Topic type:	Automatic
Single-stage retrieval?	No
Dense retrieval?	No
Used deep nn model?	Pre-trained model
Type of training:	Previous year's MS MARCO training data
Pre-processing/indexing cost:	requires expanding the original passage collection (append additional tokens to each passage); this process took about 1 hour 40 mins per 2 million passages on a single Tesla v100 16G Gpu, we use 6 gpus for this part, thus it took around 20 hours in total. After expanding the whole collection, TILDEv2 required encoding the expanded collection which costs around 40 mins per 2 million passages, we again use 6 gpus to do this. Finally, we used Anserini to indexing the encoded collection which took us around 3.5 hours on a 2018 Mac Mini machine with 3.2 GHz 6-Core Intel Core i7 CPU.
Query processing cost:	it uses a BERT tokenizer to process each query which takes only 0.1 ms per test query on cpu. This is followed by searching in our custom index (implemented using Hashtable) to compute the final scores and rerank the 1,000 passages passages retrieved by BM25, which takes around 11 ms per query on CPU.

Overall measures	
Number of topics	57
Total number retrieved	5700
Total relevant	8203
Total relevant retrieved	2091
MAP	0.2262
Mean NDCG@10	0.6022
Mean Reciprocal Rank	0.8829

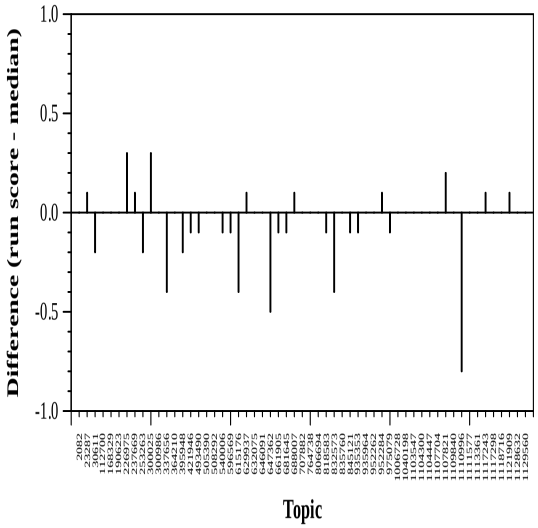
Document Level Averages	
	Precision
At 5 docs	0.8105
At 10 docs	0.7842
At 15 docs	0.7345
At 20 docs	0.6895
At 30 docs	0.6070

R-Precision	
Exact	0.2916

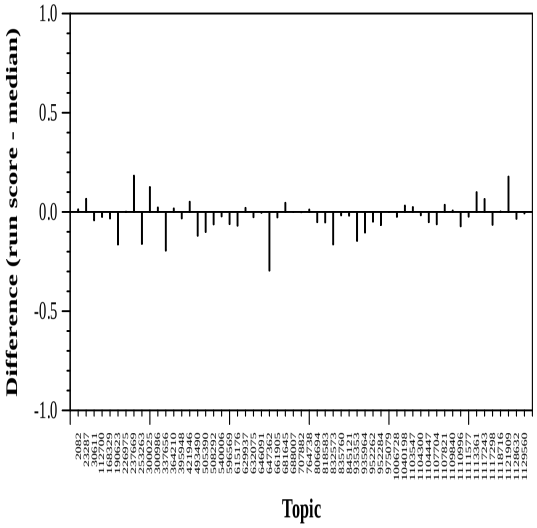




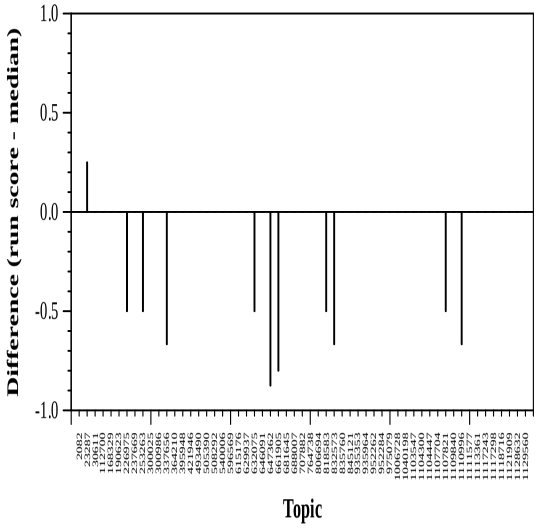
Per-topic difference from median for NDCG@10 for Document Ranking runs



Per-topic difference from median for Prec@10 for Document Ranking runs



Per-topic difference from median for Average Precision for Document Ranking runs



Per-topic difference from median for Reciprocal Rank for Document Ranking runs