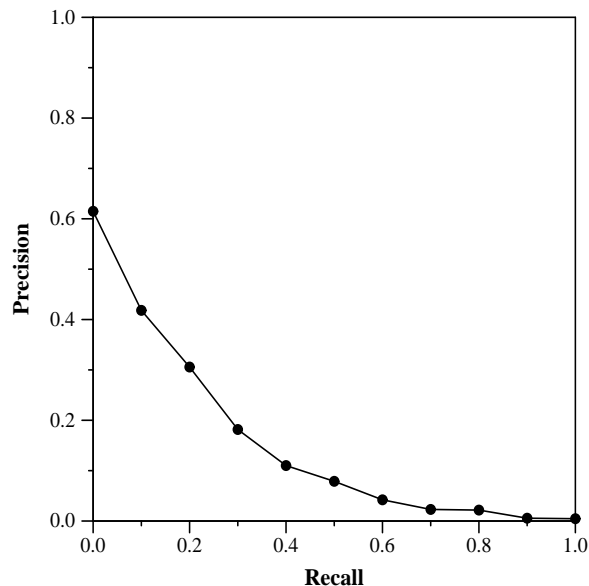
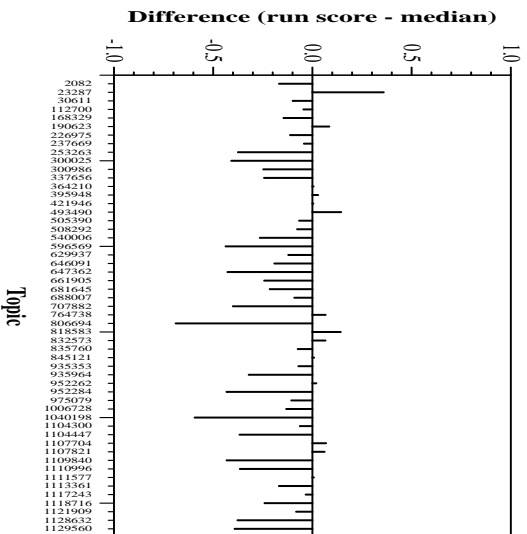


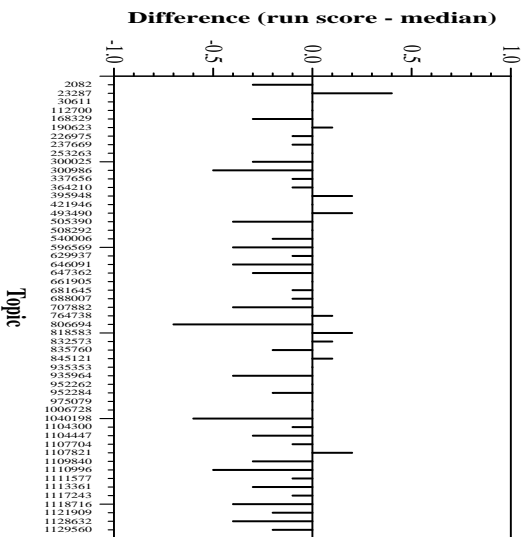
Run Description	
Run ID:	WLUPassage
Task:	Passage Ranking
Subtask:	Reranking the official top 100
Topic type:	Automatic
Single-stage retrieval?	No
Dense retrieval?	No
Used deep nn model?	Pre-trained model
Type of training:	Previous year's MS MARCO training data
Pre-processing/indexing cost:	No modifications were made to the passage collection before the query arrived. The passage corpus was downloaded, and passages were retrieved from it, and inputted into the model as-is. The only processing was done to the training set: the triples were split into query-passage pairs and a new column labelled each passage 0 or 1 depending on its relevancy to the query.
Query processing cost:	Each query (and its top-100 passages) was tested separately. For each query, two lists of length 100 were passed to the model: the same query 100 times was paired with the top-100 passages. They were converted to BERT embeddings and then inputted into the neural network model described above. The model outputted predictions for each query-passage pair, a floating point from 0 to 1. It took approximately 4 hours. Completed on 1 Tesla P100. GPU memory is 16 GB. GPU minutes = 4 hours * 60 minutes * 1 GPU = 240

Overall measures		Document Level Averages	
Number of topics	53		Precision
Total number retrieved	5300	At 5 docs	0.4000
Total relevant	3427	At 10 docs	0.3604
Total relevant retrieved	879	At 15 docs	0.3283
MAP	0.1348	At 20 docs	0.3189
Mean NDCG@10	0.4432	At 30 docs	0.3038
Mean Reciprocal Rank	0.5682	R-Precision	
		Exact	0.2031

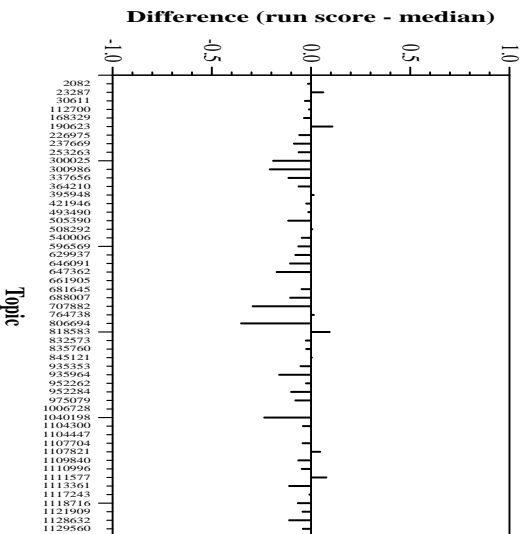




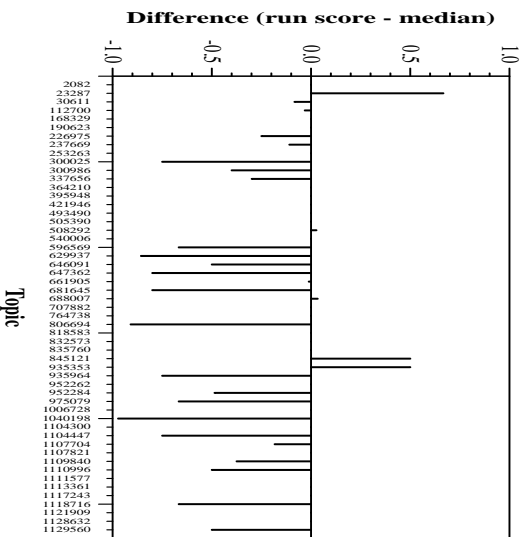
Per-topic difference from median for NDCG@10 for Passage Ranking runs



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



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



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