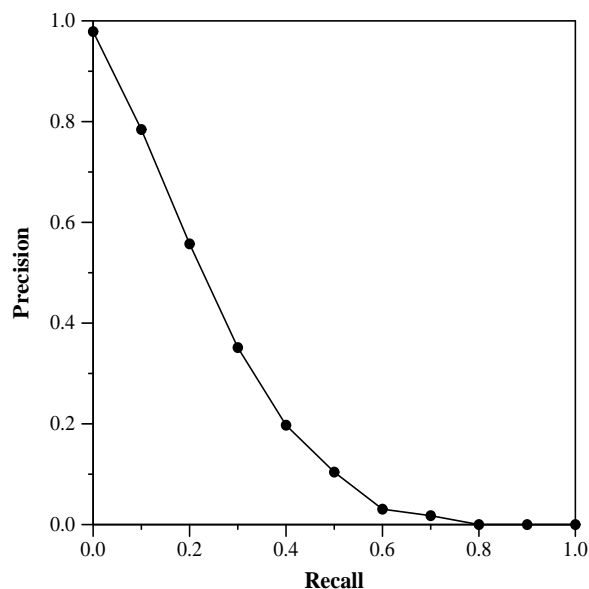
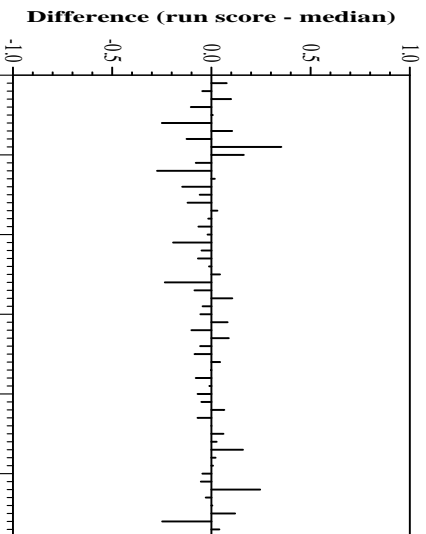


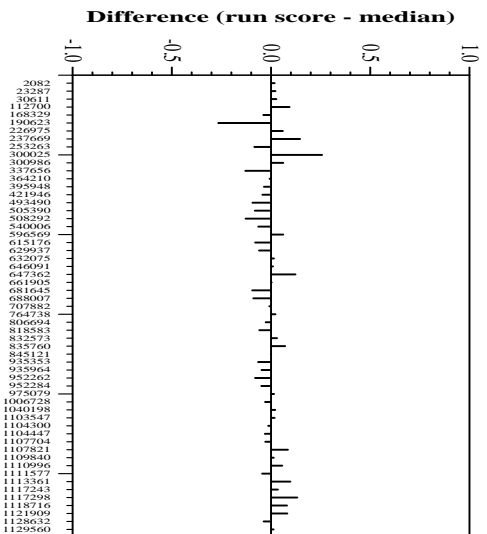
Run Description	
Run ID:	ielab-AD-uni-d
Task:	Document Ranking
Subtask:	Full ranking from the collection
Topic type:	Automatic
Single-stage retrieval?	Yes
Dense retrieval?	Yes
Used deep nn model?	Pre-trained model
Type of training:	Previous year's MS MARCO training data
Pre-processing/indexing cost:	requires tokenising the original passage collection, for which 15 minutes on average is needed per 2 million passages on a single Tesla v100 16G Gpu; we use 6 gpu for this part, which took around 3 hours for this task. After tokenisation, the STAR model needs to compute passage embeddings: this requires around 1 hour per 2 million passages, we also use 6 Tesla v100 16 gpu to conduct this task, and took overall 11 hours and 40 minutes.
Query processing cost:	For test queries, firstly it takes on average 1ms for each query in the test queries list to generate embeddings using the adore-star model, then the similarity search is about .2 ms per query per 2 million passages and therefore it takes around 84ms for the whole collection search (splited in 70 chunks). Overall, for a single query in the test collection, query processing amounts to 85 ms.

Overall measures		Document Level Averages	
Number of topics	57		Precision
Total number retrieved	5700	At 5 docs	0.8561
Total relevant	8203	At 10 docs	0.8088
Total relevant retrieved	2280	At 15 docs	0.7626
MAP	0.2492	At 20 docs	0.7114
Mean NDCG@10	0.6424	At 30 docs	0.6310
Mean Reciprocal Rank	0.9684	R-Precision	
		Exact	0.3178

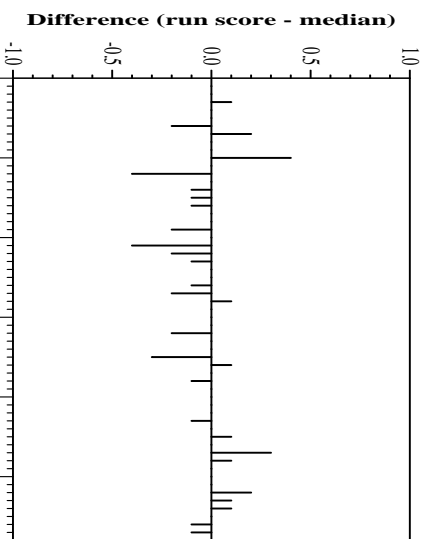




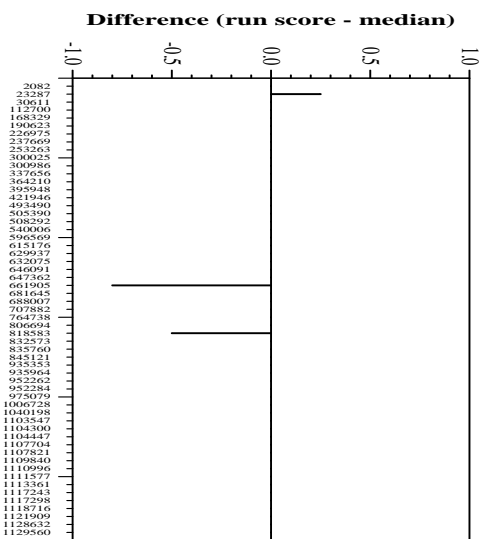
Per-topic difference from median for NDCG@10 for Document Ranking runs



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



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



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