Round 5 results — Run elhuyar_prf_nof9p submitted from Elhuyar_NLP_team

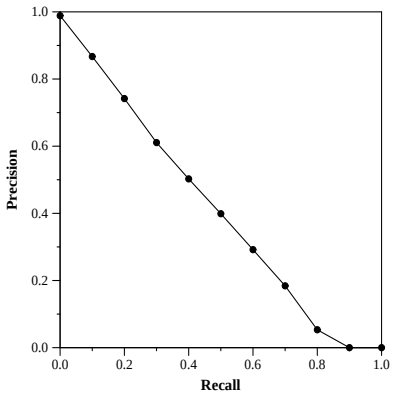
Run Description

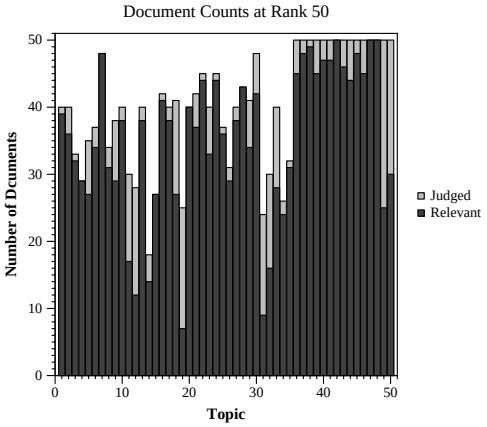
We tackle this document retrieval task in two steps: a) a first ranking and b) re-ranking. In order to obtain the first ranking of relevant documents of the collection corresponding to the queries, we use a language modeling based information retrieval approach (Ponte & Croft, 1998) including pseudo relevance feedback. For that purpose, we used the Indri search engine (Strohman, 2005), which combines Bayesian networks with language models. Full articles are indexed and titles and abstracts are expanded. When building the query, different weights are assigned to the query, question and narrative fields Then, we make a re-ranking based on BERT following a strategy similar to the one proposed by Nogueira and Cho (2019). We tuned the Clinical BERT model (Alsentzer et al., 2019) to the task of identifying relevant queries and abstracts by using a silver dataset composed of titles and their corresponding abstracts from the COVID-19 Open Research dataset and the qrels of the previous rounds. Indri and Tuned Clinical BERT scores are linearly combined and re-ranking is performed according to that new score. In this run we used a weight of 0.9 for the Clinical BERT score.

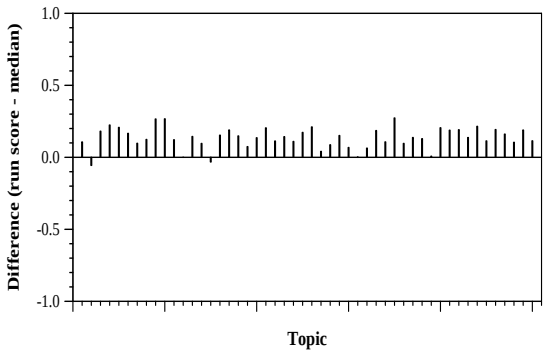
Summary Statistics	
Run ID	elhuyar_prf_nof9p
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	50
Total number retrieved	49904
Total relevant	10910
Total relevant retrieved	7525
MAP	0.4105
Mean Bpref	0.6002
Mean NDCG@20	0.7808
Mean RBP(p=0.5)	0.8524 +0.0004

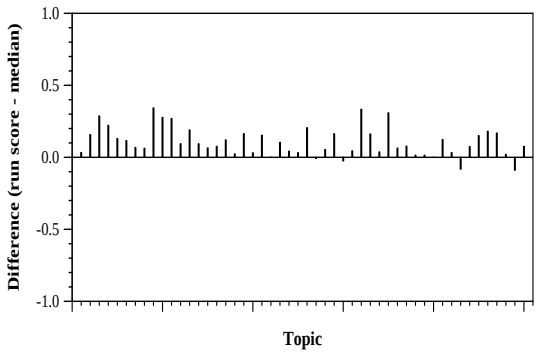
Document Level Averages	
	Precision
At 5 docs	0.9360
At 10 docs	0.9060
At 15 docs	0.8627
At 20 docs	0.8320
At 30 docs	0.7833
R-Precision	
Exact	0.4408





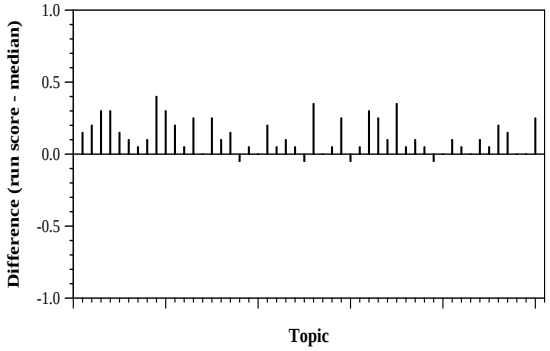


Per-topic difference from median bpref for all Round 4 runs

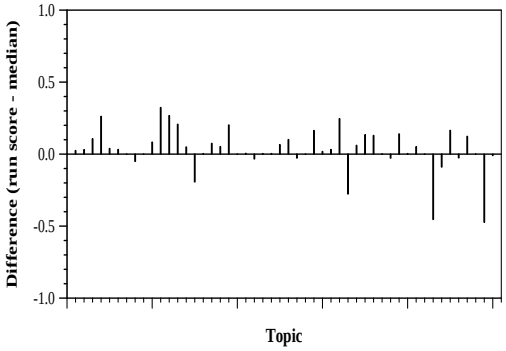


Per-topic difference from median NDCG@20 for all Round 4 runs

Round 5 results — Run elhuyar_prf_nof9p submitted from Elhuyar_NLP_team



Per-topic difference from median P@20 for all Round 4 runs



Per-topic difference from median RBP(p=0.5) for all Round 4 runs