

Summary statistics

Run ID	FW14Search100
Task	Resource selection
Resources used	Used result snippets only
External data used in run	<none>
Description	Concatenate all the sampled snippets for each resource into one large document, and run LDA topic analysis over the collection of these large documents. The topic distribution of each resource is then that of its corresponding large document. Each query is expanded by Google and then its topic distribution is inferred using the trained topic models. Finally the resources are ranked by according to the similarities of their topic distribution vectors to that of the query. In this run, the number of topics in training LDA topic models is set to 100.
Number of topics	50

Metric	Average	Median	Min	Max
nDCG@20	0.505	0.525	0.216	0.755
nDCG@10	0.425	0.446	0.123	0.692
nP@1	0.277	0.059	0.000	1.000
nP@5	0.383	0.378	0.023	0.739

