

BioCADDIE Results

Baselines

The following baselines were run using the combined, short topics on the biocaddie_all index using the baselines/ scripts to sweep parameter combinations and the mkeval.sh script to perform LOOCV for each desired metric.

Model	MAP	NDCG	P@20	NDCG@20	P@100	NDCG@100	Notes	Date
TFIDF	0.2781	0.5324	0.5405	0.4593	0.4024	0.4369	Sweep b and k1	5/17/17
Okapi	0.2894	0.5629+	0.5643	0.4927+	0.4029	0.4516	Sweep b, k1, k3	5/17/17
QL (JM)	0.2661	0.5311	0.5381	0.4853	0.3852	0.437	Sweep lambda	5/17/17
QL (Dir)	0.2976+	0.5656+	0.5643	0.5004	0.4119	0.457	Sweep mu	5/17/17
QL (TS)	0.2931+	0.5653+	0.5714	0.4827	0.4024	0.4507	Sweep mu and lambda	5/17/17

+ indicates significant improvement over TFIDF baselines ($p < 0.05$)

Results when using the combined, orig topics on the biocaddie_all index using the baselines/ scripts to sweep parameter combinations and the mkeval.sh script to perform LOOCV for each desired metric.

Model	MAP	p-value	NDCG	p-value	P@20	p-value	NDCG@20	p-value	P@100	p-value	NDCG@100	p-value	Notes	Date
TFIDF	0.1661	-	0.3749	-	0.395	-	0.3234	-	0.2505	-	0.2962	-	Sweep b and k1	5/16/17
Okapi	0.1348-	0.9938	0.3279-	0.9935	0.3225-	0.9626	0.2558-	0.9766	0.212	0.9482	0.2547-	0.9717	Sweep b, k1, k3	5/16/17
QL (JM)	0.1283-	0.98	0.3388	0.7939	0.2825	0.9267	0.2079-	0.9628	0.23	0.7278	0.2508	0.891	Sweep lambda	5/16/17
QL (Dir)	0.1447	0.9051	0.3689	0.5642	0.285-	0.9688	0.2324-	0.9676	0.2415	0.6319	0.2666	0.8402	Sweep mu	5/16/17
QL (TS)	0.1606	0.632	0.3724	0.5259	0.36	0.7977	0.304	0.738	0.2475	0.5424	0.2991	0.4605	Sweep mu and lambda	5/16/17
RM3	0.1795	0.3093	0.373	0.5156	0.4025	0.4428	0.3234	0.8438	0.247	0.5367	0.2913	0.5462	Sweep mu, fbDocs, fbTerms, and lambda	5/22/17
SDM	0.1575	0.7099	0.3962	0.3021	0.42	0.3554	0.3446	0.2989	0.2695	0.2764	0.3045	0.3995	Sweep mu, w1, w2, w3	5/22/17

***No improvement over TFIDF baselines when using original queries.

- indicates significant decrement in performance over TFIDF baselines ($p > 0.95$)

Feedback/SDM models

Green rows are runs with Garrick's framework, all others are with Craig's framework.

Model	MAP	NDCG	P@20	NDCG@20	P@100	NDCG@100	Notes	Date
QL (Dir)	0.2976+	0.5656+	0.5643	0.5004	0.4119	0.457		5/15/17
QL (Dir)	0.2853	0.5304	0.5667	0.5054	0.3986	0.4541	Sweep mu	5/15/17
RM3	0.2846	0.5381	0.5738	0.5054	0.4157	0.4710	Sweep mu, fbDocs, fbTerms, and lambda	5/15/17
RM3	0.2846	0.5395	0.5619	0.5072	0.4157	0.4710	Sweep mu, fbDocs, fbTerms, and lambda	5/16/17
RM3 (stopped)	0.2948	0.5834	0.5571	0.5004	0.3805	0.4599	Sweep mu, fbDocs, fbTerms, and lambda	5/17/17
RM3 (stopped)	0.3023	0.5483	0.5738	0.5118	0.3905	0.4626	Sweep mu, fbDocs, fbTerms, and lambda	5/16/17
Pubmed	0.2945	0.5998+	0.5190	0.4759	0.4190	0.4607	Use mu=2500 for pubmed query sweeping fbDocs, fbTerms, and lambda. Sweep mu for final retrieval.	5/15/17
Wikipedia	0.3051	0.5894+	0.5619	0.4821	0.3948	0.4613	Use mu=2500 for wikipedia query sweeping fbDocs, fbTerms, and lambda. Sweep mu for final retrieval.	5/15/17
SDM	0.2874	0.5558	0.4833	0.4706	0.4019	0.4559	Sweep mu, w1, w2, w3	5/16/17
FDM	0.293	0.5594+	0.5500	0.4960	0.3933	0.4560	Sweep mu, w1, w2, w3	5/8/17
OKAPI Exp	0.3189	0.5745	0.5857	0.4894	0.4238	0.4783	OKAPI expansion	5/17/17

+ indicates significant improvement over QL (Dir) baselines ($p < 0.05$)

Document expansion models

Model	MAP	NDCG	P@20	NDCG@20	P@100	NDCG@100	Notes	Date
QL (Dir)	0.2853	0.5304	0.5667	0.5054	0.3986	0.4541	Sweep mu	5/15/17
Pubmed QL	0.2467	0.5416	0.4147	0.3500	0.3506	0.3608	Sweep alpha	5/22/17
Pubmed RM3	0.2949	0.5806+	0.5643	0.4762	0.3819	0.4431	Sweep mu, fbDocs, fbTerms, alpha, and lambda	5/17/17

Wikipedia QL	0.2391	0.5267	0.4500	0.3727	0.3535	0.3610	Sweep alpha	5/22/17
Wikipedia RM3	0.2800	0.5795+	0.5524	0.4712	0.3724	0.4254	Sweep mu, fbDocs, fbTerms, alpha, and lambda	5/17/17
Pubmed+Wikipedia QL	0.2467	0.5416	0.4147	0.3500	0.3459	0.3608	Sweep alpha	5/22/17
Pubmed+Wikipedia RM3	0.2923	0.5725+	0.5667	0.4600	0.3933	0.4378	Sweep mu, fbDocs, fbTerms, alpha, and lambda	5/17/17

Model	MAP	NDCG	P@20	NDCG@20	P@100	NDCG@100	Notes	Date
Dir (Krovetz)	0.2959+	0.5443+	0.5905+	0.5150	0.4124+	0.4655	+ means significant improvement of Dir (unstemmed)	5/15/17
RM3 (krovetz)	0.3029	0.5734	0.5643	0.4714	0.4033	0.4753		5/17/17
Pubmed (Krovetz)	0.3285+	0.6008+	0.5667	0.4821	0.4333+	0.4627	+ means improvement of Dir (krovetz)	