Table 2: Retrieval Results (• & ∘ denote statistical significance at p-value < 0.01 & < 0.05 respectively)

Dataset	webmasters (#QAs=1299)				android (#QAs=2193)				gis (#QAs=3726)				SemEval2016 (#QAs=1366)			
t=5																
Method	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG
LASER-QA	**0.022•**	**0.101•**	**0.079•**	**0.082•**	**0.026•**	**0.127•**	**0.094•**	**0.102•**	**0.023•**	**0.111•**	**0.085•**	**0.089•**	**0.141∘**	**0.407•**	**0.266∘**	**0.265∘**
TBLM	0.009	0.043	0.031	0.034	[illegible]	0.080	0.059	0.061	0.008	0.041	0.034	0.035	0.084	0.259	0.158	0.158
TTRLM	0.012	0.056	0.030	0.033	0.022	0.101	0.061	0.067	0.016	0.079	0.048	0.052	0.067	0.201	0.134	0.139
AENN	0.008	0.035	0.023	0.026	0.007	0.033	0.012	0.012	0.005	0.026	0.016	0.017	0.030	0.148	0.060	0.059
t=10																
Method	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG
LASER-QA	**0.013•**	**0.116•**	**0.080•**	**0.088•**	**0.015•**	**0.148•**	**0.097•**	**0.119•**	**0.013•**	**0.123•**	**0.086•**	**0.095•**	**0.125•**	**0.556•**	**0.291•**	**0.292•**
TBLM	0.006	0.050	0.031	0.036	0.010	0.089	0.060	0.066	0.005	0.046	0.034	0.037	0.070	0.296	0.156	0.163
TTRLM	0.010	0.093	0.036	0.043	0.014	0.124	0.065	0.076	0.010	0.098	0.050	0.060	0.069	0.383	0.152	0.160
AENN	0.005	0.043	0.024	0.029	0.005	0.047	0.014	0.018	0.003	0.030	0.016	0.019	0.036	0.259	0.069	0.078
t=20																
Method	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG
LASER-QA	**0.007**	**0.121**	**0.080•**	**0.093•**	**0.008**	**0.155**	**0.097•**	**0.116•**	**0.007**	**0.126**	**0.087•**	**0.100•**	**0.079**	**0.605**	**0.296•**	**0.327•**
TBLM	0.003	0.050	0.031	0.038	0.006	0.100	0.060	0.070	0.003	0.052	0.034	0.038	0.051	0.433	0.157	0.175
TTRLM	0.006	0.118	0.037	0.052	**0.008**	0.149	0.066	0.085	0.006	0.119	0.052	0.067	0.059	0.519	0.149	0.182
AENN	0.003	0.051	0.025	0.032	0.003	0.066	0.015	0.023	0.002	0.030	0.017	0.021	0.023	0.321	0.074	0.104
t=50																
Method	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG	Prec	SR	MAP	NDCG
LASER-QA	**0.003**	0.125	**0.081•**	**0.096•**	0.003	0.166	**0.097•**	**0.121•**	0.002	0.129	**0.087•**	**0.103•**	0.033	0.636	**0.295•**	**0.359•**
TBLM	0.001	0.056	0.031	0.040	0.003	0.142	0.060	0.074	0.001	0.060	0.035	0.040	0.026	0.646	0.151	0.187
TTRLM	**0.003**	**0.145•**	0.038	0.061	**0.004**	0.177	0.066	0.093	**0.003•**	**0.132•**	0.052	0.074	**0.039**	**0.667**	0.140	0.214
AENN	0.002	0.070	0.025	0.035	0.002	0.088	0.015	0.028	0.001	0.065	0.017	0.025	0.016	0.407	0.066	0.107

Dataset	#QAs	t=5 Prec	t=5 SR	t=5 MAP	t=5 NDCG	t=50 Prec	t=50 SR	t=50 MAP	t=50 NDCG
stats	4004	**0.016•**	**0.076•**	**0.057•**	**0.060•**	0.002	0.096	**0.058•**	**0.071•**
programmers	4107	**0.020•**	**0.096•**	**0.068•**	**0.075•**	0.002	0.115	**0.069•**	**0.088•**
wordpress	4744	**0.019•**	**0.091•**	**0.069•**	**0.074•**	0.002	0.112	**0.070•**	**0.085•**
physics	5044	**0.025•**	**0.120•**	**0.088•**	**0.094•**	0.003	0.148	**0.090•**	**0.111•**
mathematica	5084	**0.018•**	**0.087•**	**0.067•**	**0.072•**	0.002	0.116	**0.069•**	**0.084•**
unix	5330	**0.023•**	**0.115•**	**0.089•**	**0.094•**	0.003	0.137	**0.091•**	**0.107•**
gaming	6398	**0.034•**	**0.166•**	**0.130•**	**0.137•**	0.004	0.189	**0.132•**	**0.155•**
english	6668	**0.024•**	**0.115•**	**0.090•**	**0.095•**	**0.003**	**0.130**	**0.092•**	**0.107•**

Table 3: LASER-QA Results (Boldfacing and Statistical Significance indications from comparison with TopicTRLM and TBLM) over Larger Categories in CQADupStack

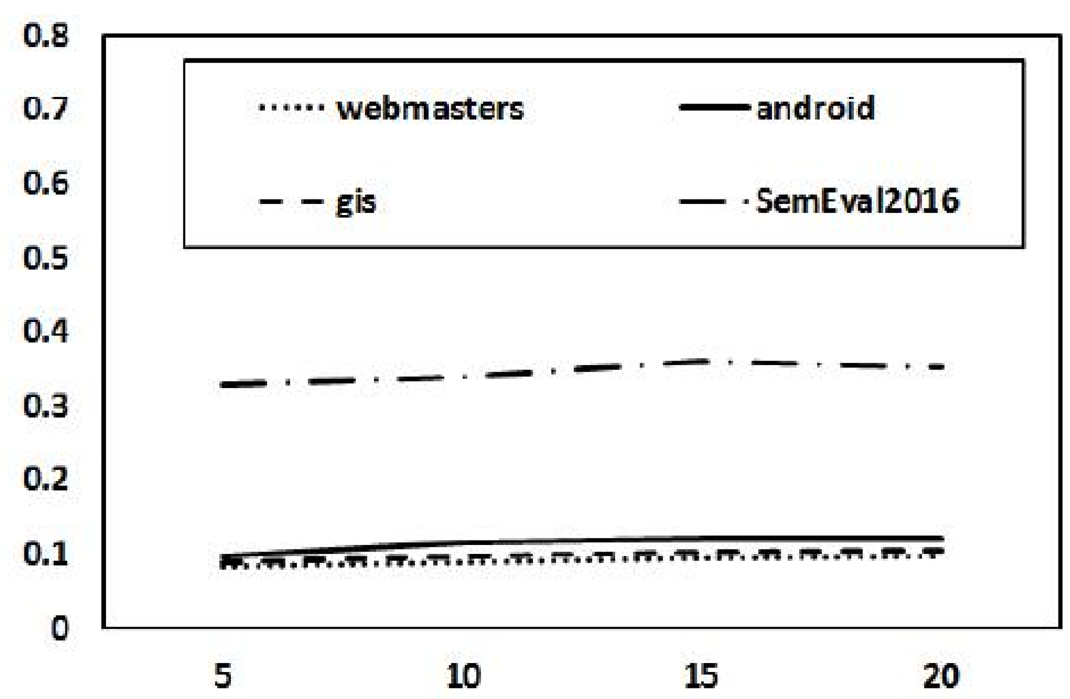

Figure 1: NDCG (Y-axis) v/s. k

creases way beyond the training neighborhood size (i.e., 15), LASER-QA is seen to deteriorate gracefully (as expected).

- LASER-QA performance peaks on rank-aware metrics such as MAP and NDCG (even at $t = 50$), indicating it's high effectiveness in producing relevant results at the top.

These observations underline the effectiveness of LASER-QA as a CQA retrieval method. It may be noted that LASER-QA uses compact representations ($d < 2000$), as compared to vocabulary space representations that are typically ≥ 5000.

Trends at High t: The performance trends at high values of t are explained by the usage of the local neighborhood (of size k) in LASER-QA latent space learning. Going down the result list much beyond k reveals expected, but graceful, decline in accuracy. For automated processing scenarios that necessitate large t, a correspondingly high k may be used in learning. It is notable that LASER-QA's focus on local neighborhood manifests as enhanced accuracy at the top of the result set.

LASER-QA Analysis on Larger CQADupStack Datasets: Owing to scalability issues of AENN that disallows a full evaluation over larger categories in CQADupStack, we present LASER-QA results over them in Table 3 to illustrate the consistency in trends. Boldfacing and statistical significance have the same semantics as earlier, with the comparison performed against only TopicTRLM and TBLM.

5.3 LASER-QA Parameter Analysis

We now analyze the NDCG trends (NDCG being the most popular IR metric) across LASER-QA parameters, i.e., k, α and d, varying each one separately keeping the default choice for others.

- **Varying k:** Figure 1 plots NDCG against values of k from $\{5, 10, 15, 20\}$. As may be seen, the accuracy is seen to improve with increasing k in the lower ranges, while sat-