

TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results

Group: (SetuServ) SetuServ

Run ID: SSEC3incl

Run type: Secondary

Description of run:

Post Error Correction (likely errors are identified using M/L), counting “not sures” as relevant documents

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	22	1974	55	5	0.804	0.973	0.027	0.196	0.077	0.968
416	1235	45	41	1180	10	4	0.902	0.991	0.009	0.098	0.030	0.965
417	2992	75	60	2808	109	15	0.796	0.962	0.038	0.204	0.091	0.890
420	1136	37	34	1050	49	3	0.908	0.955	0.045	0.092	0.065	0.960
427	1528	37	31	1477	14	6	0.829	0.990	0.010	0.171	0.043	0.956
432	2503	22	9	2408	73	13	0.413	0.970	0.030	0.587	0.172	0.712
438	1798	162	153	1497	139	9	0.942	0.915	0.085	0.058	0.071	0.951
445	1404	60	50	1325	19	10	0.828	0.986	0.014	0.172	0.052	0.935
446	2020	156	98	1809	55	58	0.627	0.970	0.030	0.373	0.119	0.834
447	1588	16	15	1570	2	1	0.912	0.998	0.002	0.088	0.012	0.968
Average	1826.000	63.700	51.300	1709.800	52.500	12.400	0.796	0.971	0.029	0.204	0.073	0.914

Table 1: This table shows per-topic statistics and overall averages for the run SSEC3incl. The topics are 10 randomly selected topics from the TREC 8 ad-hoc task. A relevant document is positive and a non-relevant document is negative. The true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts are based on an adjudicated set of relevance judgments that differs from the original TREC-8 ad-hoc grels. The true positive rate (TPR), false positive rate (FPR), true negative rate (TNR), and the false negative rate (FNR) are all smoothed values. Details of the computation of the logistic average misclassification (LAM) rate and the area under the curve (AUC) are given in the track overview paper. Some runs did not report a probability of relevance and thus will have NA for their AUC score.

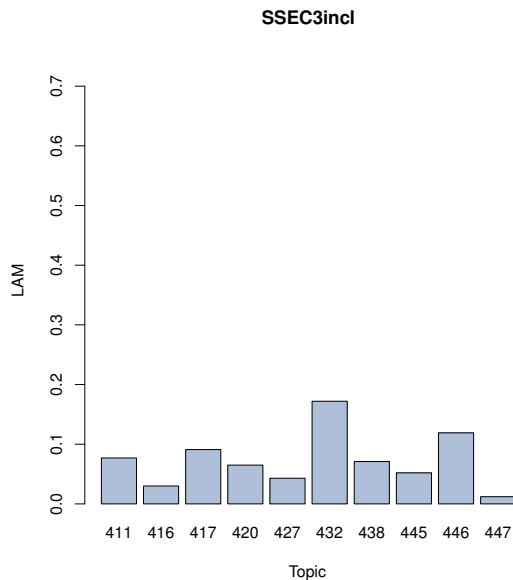


Figure 1: SSEC3incl LAM

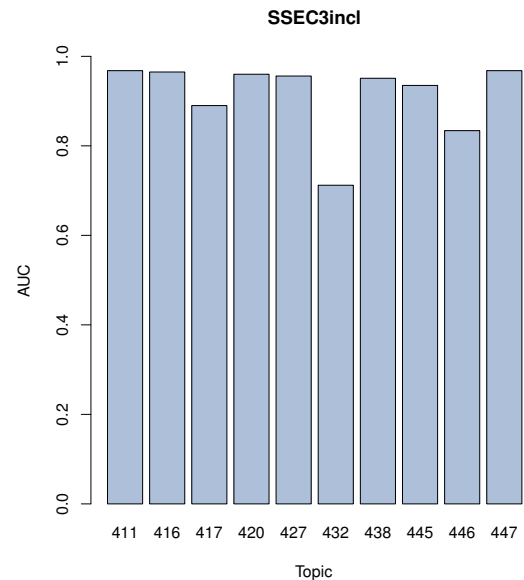


Figure 2: SSEC3incl AUC