TREC 2012 Crowdsourcing Track, Text Relevance Assessing Task (TRAT) results
Group: (SetuServ) SetuServ
Run ID: SSEC3excl

Run type: Secondary
Description of run:
Post Error Correction (likely errors are identified using M/L), counting "not sures" as irrelevant documents

Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	21	1992	37	6	0.768	0.982	0.018	0.232	0.070	0.968
416	1235	45	40	1187	3	5	0.880	0.997	0.003	0.120	0.020	0.965
417	2992	75	46	2864	53	29	0.612	0.982	0.018	0.388	0.098	0.890
420	1136	37	33	1061	38	4	0.882	0.965	0.035	0.118	0.065	0.960
427	1528	37	31	1482	9	6	0.829	0.994	0.006	0.171	0.035	0.956
432	2503	22	9	2427	54	13	0.413	0.978	0.022	0.587	0.152	0.712
438	1798	162	151	1548	88	11	0.929	0.946	0.054	0.071	0.062	0.951
445	1404	60	43	1329	15	17	0.713	0.988	0.012	0.287	0.064	0.935
446	2020	156	91	1829	35	65	0.583	0.981	0.019	0.417	0.105	0.834
447	1588	16	7	1572	0	9	0.441	1.000	0.000	0.559	0.020	0.968
Average	1826.000	63.700	47.200	1729.100	33.200	16.500	0.705	0.981	0.019	0.295	0.069	0.914

Table 1: This table shows per-topic statistics and overall averages for the run SSEC3excl. 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 qrels. 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.
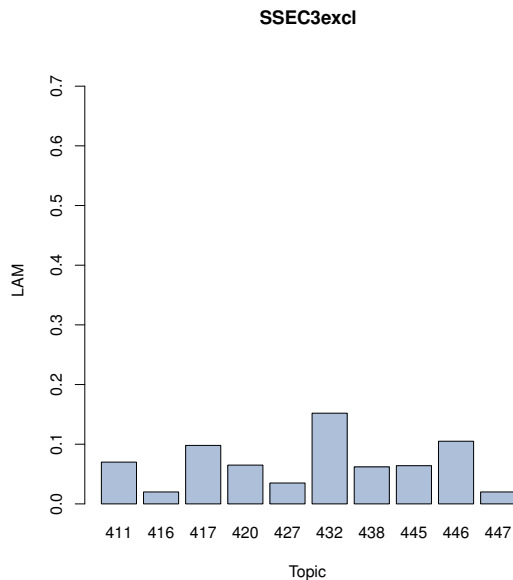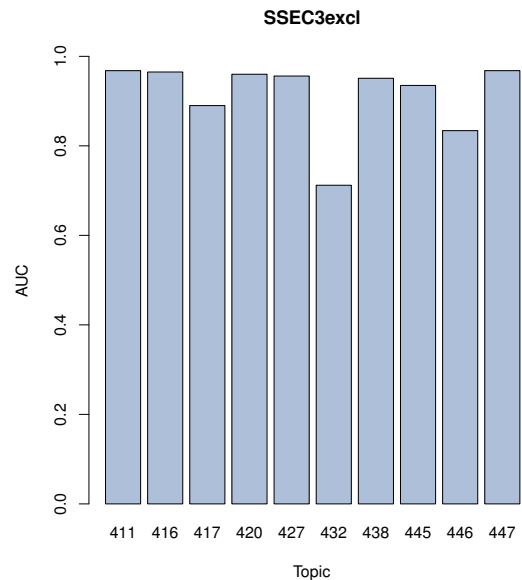

Figure 1: SSEC3excl LAM

Figure 2: SSEC3excl AUC