

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

Group: (SetuServ) SetuServ

Run ID: SSML2pct

Run type: Secondary

Description of run:

Crowdsourced using Skierarchy approach for pairs with machine learning score of greater than 2%. Remaining documents are marked as not relevant

## Results

Topic	#Docs	#Rel	TP	TN	FP	FN	TPR	TNR	FPR	FNR	LAM	AUC
411	2056	27	24	1968	61	3	0.875	0.970	0.030	0.125	0.063	0.933
416	1235	45	41	1177	13	4	0.902	0.989	0.011	0.098	0.034	0.953
417	2992	75	58	2826	91	17	0.770	0.969	0.031	0.230	0.090	0.880
420	1136	37	34	1055	44	3	0.908	0.960	0.040	0.092	0.061	0.961
427	1528	37	26	1481	10	11	0.697	0.993	0.007	0.303	0.053	0.889
432	2503	22	9	2434	47	13	0.413	0.981	0.019	0.587	0.143	0.717
438	1798	162	141	1514	122	21	0.868	0.925	0.075	0.132	0.100	0.916
445	1404	60	44	1326	18	16	0.730	0.986	0.014	0.270	0.067	0.876
446	2020	156	90	1812	52	66	0.576	0.972	0.028	0.424	0.127	0.811
447	1588	16	14	1571	1	2	0.853	0.999	0.001	0.147	0.013	0.969
Average	1826.000	63.700	48.100	1716.400	45.900	15.600	0.759	0.974	0.026	0.241	0.075	0.891

Table 1: This table shows per-topic statistics and overall averages for the run SSML2pct. 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.

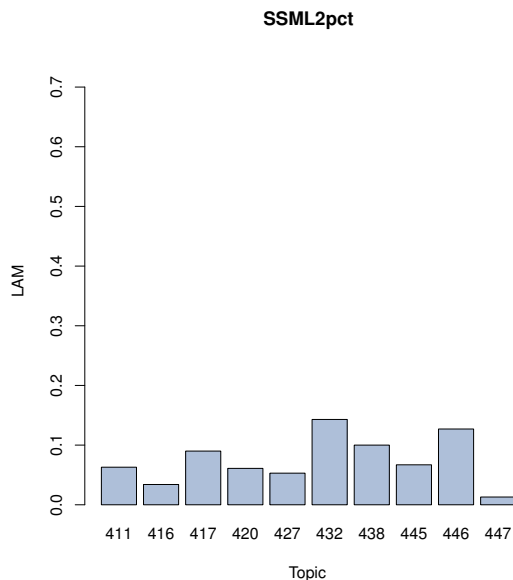


Figure 1: SSML2pct LAM

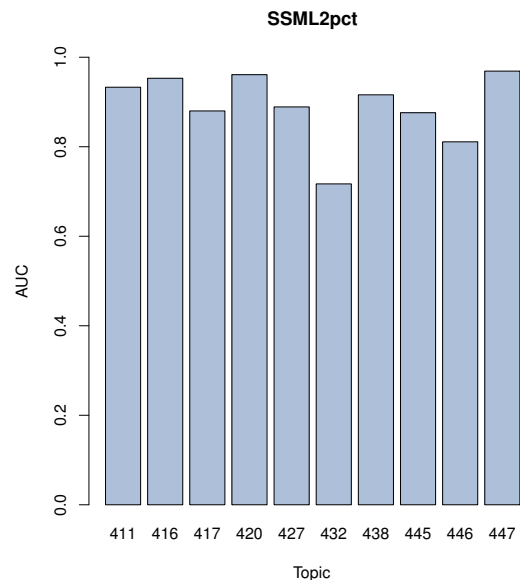


Figure 2: SSML2pct AUC