Mean Scores: TREC 2008 Legal Track, Relevance Feedback Task

(Revised Feb 16/09 to make the table captions consistent with the overview paper.)

The tables list the mean scores for each of the 29 submitted runs for the Relevance Feedback task and the 2 additional reference runs. Note that residual evaluation was used for all measures (i.e., documents judged in previous years were removed before evaluating). The following glossary explains the codes used in the tables.

"Fields": The topic fields used by the run: 'b' Boolean query (final negotiated), 'C' complaint, 'd' defendant Boolean (initial proposal), 'i' instructions and definitions, 'p' plaintiff Boolean (rejoinder query), 'r' request text, 'v' B value, 'm' metadata fields were indexed, 'B' reference Boolean run was used, 'M' manual processing was involved, 'F' feedback run (old relevance assessments were used, applicable to RF task only).

"Ret_r": The Average (Residual) Number of Documents Retrieved per Topic.

"K_r": The Average residual K value.

"P@K_r" and "R@K_r": Estimated Precision and Recall at Depth K_r.

"$F_1\text{@K}_r$": Estimated F_1 at Depth K_r.

"$F_1\text{@R}_r$": Estimated F_1 at Depth R_r (where R_r is the estimated number of residual relevant documents).

"S1J": Success of the First Judged Document.

"P5": Estimated Precision at Depth 5.

"R@B_r": Estimated Recall at Depth B_r.

"R@ret_r": Estimated Recall of the full retrieval set.

"K_{hr}": Residual K value when just counting Highly relevant documents as relevant.

"R_{hr}": Estimated number of residual Highly relevant documents.

Table 1 counts all relevant documents as relevant (averaged over 12 topics). Table 2 just counts highly relevant documents as relevant (averaged over 9 topics).

Rank-based measures are omitted for the (unranked) reference Boolean run.

Run	Fields	Ret$_r$	K$_r$	(P@K$_r$, R@K$_r$)	F_1@K$_r$	F_1@R$_r$	S1J, P5	R@B$_r$, R@ret$_r$
refRF08B	bvmBM	3488	3488	(0.367, 0.228)	**0.142**			0.228, 0.228
otRF08fb	bvmBM	3488	3488	(0.367, 0.228)	**0.142**	0.151	9/12, 0.656	0.228, 0.228
otRF08rvl	rmM	100000	87738	(0.126, **0.607**)	0.134	0.185	6/12, 0.450	0.096, 0.640
UMKCTL08RF6	F-bBM	100000	3476	(0.367, 0.212)	0.131	**0.309**	8/12, 0.667	0.212, 0.690
otRF08fv	bmM	100000	70255	(0.099, 0.495)	0.119	0.202	8/12, 0.533	0.103, 0.596
otRF08frw	brmBM	100000	39726	(0.140, 0.549)	0.116	0.307	**10/12**, **0.800**	0.236, 0.659
UMKCTL08RF3	F-bM	100000	859	(**0.377**, 0.142)	0.109	0.239	7/12, 0.583	0.154, 0.662
IowaSL08RF3B		50663	2442	(0.249, 0.207)	0.092	0.209	8/12, 0.650	0.199, 0.545
UMKCTL08RF2	F-bM	100000	880	(0.371, 0.108)	0.091	0.206	8/12, 0.667	0.120, 0.626
UMKCTL08RF1	F-bM	100000	875	(0.327, 0.124)	0.088	0.225	8/12, 0.629	0.148, 0.669
otRF08fbF	F-bmBM	100000	36620	(0.118, 0.392)	0.084	0.261	8/12, 0.633	**0.261**, 0.472
UMKCTL08RF5	F-bBM	100000	3519	(0.348, 0.149)	0.083	0.237	8/12, 0.667	0.144, 0.670
otRF08fbFR	F-bmBM	100000	8776	(0.266, 0.261)	0.082	0.261	8/12, 0.633	**0.261**, 0.472
IowaSL08RF1B		75148	2442	(0.287, 0.150)	0.082	0.204	8/12, 0.589	0.135, **0.694**
otRF08F	F-mM	100000	36716	(0.109, 0.327)	0.075	0.156	7/12, 0.600	0.118, 0.446
SabL08rf1	F-bdporm	99609	3547	(0.318, 0.145)	0.075	0.248	9/12, 0.717	0.145, 0.606
IowaSL08RF3A	F	50663	2442	(0.227, 0.148)	0.062	0.150	3/12, 0.383	0.163, 0.545
SabL08rfbase	bdporm	100000	3517	(0.313, 0.111)	0.060	0.167	5/12, 0.422	0.110, 0.579
IowaSL08RFTr	F-bdpr	100000	2302	(0.214, 0.116)	0.058	0.160	6/12, 0.650	0.118, 0.617
IowaSL08RF1A	F	75148	2354	(0.150, 0.121)	0.058	0.197	5/12, 0.367	0.132, 0.689
IowaSL08RF2B		15944	2288	(0.231, 0.104)	0.053	0.099	6/12, 0.517	0.085, 0.300
IowaSL08RF2A	F	15944	2344	(0.209, 0.098)	0.051	0.084	4/12, 0.283	0.080, 0.300
UCBM25T10Th5	F-r	100000	21406	(0.257, 0.086)	0.048	0.126	3/12, 0.392	0.069, 0.395
otRF08FR	F-mM	100000	8826	(0.203, 0.129)	0.047	0.156	7/12, 0.600	0.118, 0.446
UCPwrT10Th5	F-r	100000	13233	(0.172, 0.113)	0.036	0.092	3/12, 0.267	0.066, 0.445
IowaSL08RF2C	F	15944	2351	(0.176, 0.086)	0.034	0.091	4/12, 0.333	0.083, 0.300
UCBM25T5Th5	F-r	100000	6517	(0.240, 0.037)	0.021	0.121	6/12, 0.433	0.070, 0.377
UCPwrT5Th5	F-r	100000	11880	(0.110, 0.101)	0.020	0.079	2/12, 0.267	0.062, 0.442
UCRFPwrBL	r	100000	1547	(0.113, 0.034)	0.017	0.089	1/12, 0.117	0.054, 0.420
UCRFBM25BL	r	100000	564	(0.200, 0.014)	0.013	0.128	3/12, 0.283	0.074, 0.432
randomRF08		100000	20000	(0.005, 0.002)	0.003	0.002	0/12, 0.000	0.002, 0.018

Table 1: Mean scores for submitted Relevance Feedback task runs, using All Relevant documents.

Run	Fields	Ret$_r$	K$_{hr}$	(P@K$_{hr}$, R@K$_{hr}$)	F_1@K$_{hr}$	F_1@R$_{hr}$	S1J, P5	R@B$_r$, R@ret$_r$
UMKCTL08RF2	F-bM	100000	879	(0.109, 0.259)	**0.136**	0.126	3/9, 0.222	0.181, 0.607
otRF08frw	brmBM	100000	2392	(0.124, 0.336)	0.129	0.145	2/9, 0.289	0.347, 0.647
UMKCTL08RF3	F-bM	100000	857	(0.121, 0.269)	0.128	0.158	2/9, 0.222	0.237, 0.699
UMKCTL08RF6	F-bBM	100000	3857	(0.117, 0.358)	0.124	**0.204**	4/9, 0.311	0.359, 0.694
UMKCTL08RF1	F-bM	100000	872	(0.098, 0.286)	0.121	0.157	1/9, 0.222	0.250, 0.693
refRF08B	bvmBM	3870	3870	(0.112, 0.336)	0.118			0.336, 0.336
otRF08fb	bvmBM	3870	3870	(0.112, 0.336)	0.118	0.148	2/9, 0.259	0.336, 0.336
UMKCTL08RF5	F-bBM	100000	3894	(0.083, 0.239)	0.092	0.146	3/9, 0.222	0.239, 0.694
IowaSL08RF3B		55146	1035	(0.105, 0.284)	0.084	0.123	**6/9**, 0.378	0.346, 0.540
IowaSL08RFTr	F-bdpr	100000	923	(**0.146**, 0.157)	0.076	0.080	4/9, **0.422**	0.159, 0.583
IowaSL08RF1A	F	79742	946	(0.057, 0.187)	0.066	0.093	1/9, 0.111	0.243, 0.794
IowaSL08RF1B		79742	1035	(0.094, 0.169)	0.065	0.104	**6/9**, 0.393	0.223, **0.795**
UCBM25T10Th5	F-r	100000	18905	(0.052, 0.257)	0.054	0.096	0/9, 0.111	0.117, 0.607
otRF08fbFR	F-bmBM	100000	10150	(0.096, 0.400)	0.050	0.202	3/9, 0.267	**0.367**, 0.661
SabL08rfbase	bdporm	100000	3896	(0.064, 0.146)	0.049	0.053	1/9, 0.148	0.144, 0.686
otRF08fbF	F-bmBM	100000	15305	(0.032, **0.516**)	0.044	0.202	3/9, 0.267	**0.367**, 0.661
IowaSL08RF2B		16229	912	(0.096, 0.120)	0.044	0.093	4/9, 0.222	0.147, 0.459
SabL08rf1	F-bdporm	99628	3926	(0.034, 0.176)	0.042	0.103	1/9, 0.200	0.175, 0.703
IowaSL08RF3A	F	55146	1035	(0.040, 0.204)	0.037	0.095	2/9, 0.244	0.287, 0.540
UCPwrT10Th5	F-r	100000	8934	(0.032, 0.174)	0.032	0.041	0/9, 0.000	0.186, 0.728
otRF08rvl	rmM	100000	39243	(0.018, 0.418)	0.029	0.036	3/9, 0.200	0.133, 0.718
otRF08FR	F-mM	100000	10213	(0.016, 0.311)	0.028	0.089	2/9, 0.244	0.237, 0.658
IowaSL08RF2A	F	16229	956	(0.029, 0.110)	0.027	0.046	1/9, 0.089	0.126, 0.459
IowaSL08RF2C	F	16229	963	(0.020, 0.115)	0.025	0.078	1/9, 0.133	0.131, 0.459
otRF08fv	bmM	100000	22248	(0.018, 0.371)	0.025	0.085	1/9, 0.156	0.185, 0.686
UCPwrT5Th5	F-r	100000	4557	(0.029, 0.044)	0.020	0.031	0/9, 0.067	0.095, 0.737
otRF08F	F-mM	100000	15390	(0.009, 0.444)	0.015	0.089	2/9, 0.244	0.237, 0.658
UCBM25T5Th5	F-r	100000	3980	(0.022, 0.025)	0.015	0.079	0/9, 0.133	0.109, 0.609
UCRFPwrBL	r	100000	265	(0.021, 0.008)	0.007	0.019	0/9, 0.022	0.069, 0.505
UCRFBM25BL	r	100000	323	(0.051, 0.013)	0.006	0.026	1/9, 0.100	0.106, 0.568
randomRF08		100000	10000	(0.000, 0.000)	0.000	0.000	0/9, 0.000	0.000, 0.000

Table 2: Mean scores for submitted Relevance Feedback task runs, using only Highly Relevant documents.