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

“Kr”: The Average residual K value.

“P@Kr” and “R@Kr”: Estimated Precision and Recall at Depth Kr.

“F1@Kr”: Estimated F1 at Depth Kr.

“F1@Rr”: Estimated F1 at Depth Rr (where Rr is the estimated number of residual relevant documents).

“S1J”: Success of the First Judged Document.

“P5”: Estimated Precision at Depth 5.

“R@Br”: Estimated Recall at Depth Br.

“R@ret”: Estimated Recall of the full retrieval set.

“Khhr”: Residual K value when just counting Highly relevant documents as relevant.

“Rhr”: 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.