

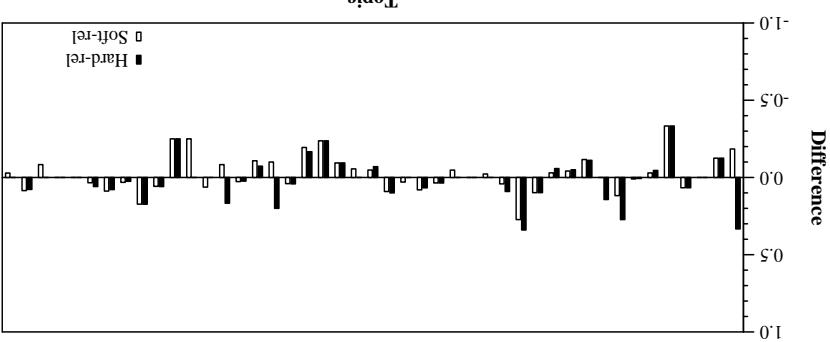
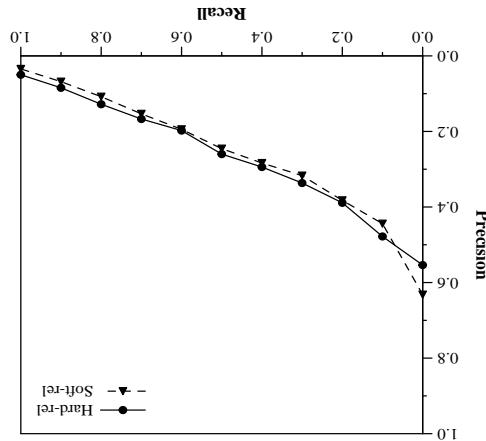
Document-level Evaluation (45 topics)

Document-level Averages		Precision		R-Precision	
Precision	Hard-rel Soft-rel	Total retrieved	44764 44764	Total relevant	0.2507 0.2411
At 5 docs	0.3467 0.3911	At 10 docs	0.3244 0.3756	At 15 docs	0.3096 0.3615
At 20 docs	0.2856 0.3367	At 30 docs	0.2704 0.3193	At 50 docs	0.2307 0.2661
At 100 docs	0.1889 0.2307	At 200 docs	0.1366 0.1661	At 300 docs	0.0960 0.1166
At 500 docs	0.0784 0.0960	At 1000 docs	0.0473 0.0581		

The figure consists of two vertically stacked plots sharing a common x-axis representing Recall from 0.0 to 1.0.

- Precision-Recall Plot:** The top plot shows Precision on the y-axis (0.0 to 1.0). It contains two data series: "Hard-rel" (solid line with circles) and "Soft-rel" (dashed line with inverted triangles). Both series show a decreasing trend as recall increases, with the Soft-rel model generally achieving higher precision than the Hard-rel model at higher recall levels.
- Topic Difference Plot:** The bottom plot shows the "Difference" on the y-axis (from -1.0 to 1.0). The legend indicates "Hard-rel" (black bar) and "Soft-rel" (white bar). The plot shows numerous small bars scattered across the recall range, with a few larger bars appearing around 0.15, 0.35, and 0.75 recall.

Document Level Averages	Precision	Hard-rel	Soft-rel	At 5 docs	At 10 docs	At 15 docs	At 20 docs	At 30 docs	At 100 docs	At 200 docs	At 300 docs	At 500 docs	At 1000 docs	R-Precision	
0.3467	0.3911	0.3244	0.3756	0.3096	0.3615	0.2856	0.3367	0.2704	0.3193	0.1889	0.2307	0.1366	0.1661	0.0784	0.0581
0.4111	0.4555	0.4000	0.4444	0.3556	0.4000	0.3000	0.3556	0.2704	0.3193	0.1889	0.2307	0.1366	0.1661	0.0784	0.0473
0.4667	0.5000	0.4111	0.4667	0.3615	0.4000	0.3000	0.3615	0.2704	0.3193	0.1889	0.2307	0.1366	0.1661	0.0784	0.0473
0.4944	0.5278	0.4222	0.4944	0.3756	0.4000	0.3000	0.3756	0.2856	0.3367	0.1990	0.2444	0.1444	0.1778	0.0960	0.0581
0.5000	0.5333	0.4286	0.5000	0.3889	0.4000	0.3000	0.3889	0.2917	0.3417	0.2000	0.2500	0.1500	0.1833	0.1000	0.0541



Passage-level Evaluation (25 topics)

Passage-based Eval Averages	Precision F Score	Charcter-based Eval Averages	
At 5 passages	0.1933 0.0638	Bpref@6000Chars 0.211 Bpref@12000Chars 0.200 Bpref@24000Chars 0.165 Bpref@48000Chars 0.1383 Bpref@96000Chars 0.1169 Bpref@192000Chars 0.10638	
At 10 passages	0.1659 0.0853	At 15 passages	0.1341 0.0945
At 20 passages	0.1411 0.0964	At 30 passages	0.1219 0.1037
At 30 passages	0.1215 0.215	At 50 passages	0.1079 0.0991
At 50 passages	0.176 0.176	At 100 passages	0.0841 0.0990
At 100 passages	0.134 0.134	Char-R-Prec	0.0717

Difference from median passage-based R-precision

Top 25 topics

RGUEI
non-baseline run, used: metadata; metadata types included Family, Genre, Geography; manual run