

Document-level Evaluation (45 topics)

Document Level Average	Precision	Hard-rel	Soft-rel	MAP
At 5 docs	0.3244	0.4222	0.2787	0.2794
At 10 docs	0.3178	0.3978	0.3007	0.3748
At 15 docs	0.3022	0.3700	0.3026	0.3747
At 20 docs	0.2793	0.3452	0.2603	0.2261
At 30 docs	0.2167	0.2113	0.1713	0.1713
At 50 docs	0.1467	0.1154	0.1154	0.1154
At 100 docs	0.1154	0.1467	0.1467	0.1467
At 200 docs	0.0852	0.1154	0.1154	0.1154
At 500 docs	0.0672	0.0852	0.0852	0.0852
At 1000 docs	0.0401	0.0502	0.0401	0.0401
Total relevant	0.3183	0.2867	0.2787	0.2794

The figure displays a Precision-Recall curve comparing two retrieval models: Hard-rel (solid line with circles) and Soft-rel (dashed line with inverted triangles). The x-axis represents Recall from 0.0 to 1.0, and the y-axis represents Precision from 0.0 to 1.0. Both curves start at a precision of approximately 0.95 at 0.0 recall and decrease as recall increases. The Soft-rel model consistently outperforms the Hard-rel model across all recall levels.

This bar chart shows the difference from the median R-precision for each topic level. The y-axis ranges from -1.0 to 1.0. The legend indicates that white bars represent Hard-rel and black bars represent Soft-rel. Most topics show a positive difference, with Soft-rel generally having higher R-precision than Hard-rel.

Passage-level Evaluation (25 topics)

Passage-based Eval Averages		Character-based Eval Averages	
	Precision F Score		R-Precision
At 5 passages	0.2657 0.0839	Bpref@6000Chars	0.217
At 10 passages	0.1976 0.0809	Bpref@12000Chars	0.185
At 15 passages	0.1939 0.0920	Bpref@24000Chars	0.186
At 20 passages	0.1869 0.0926	Bpref@60000Chars	0.226
At 30 passages	0.1787 0.1035	Bpref@120000Chars	0.209
At 50 passages	0.1677 0.1160	Bpref@240000Chars	0.224
At 100 passages	0.1344 0.1250	Char-R-Prec	0.181
		Difference	

Topic

Difference

Topic

Difference

Run ID	Run Description
UMassF	non-baseline run, used: passages, automatic run using title+desc