Terabyte Track results — Microsoft Research Asia

Summary Statistics	
Run ID	MSRAt1
Run Type	automatic
Fields Used	title
Run Description	NoLinksUsed-NoAnchorTextUsed-DocStructUsed
Time to index	700 min.
Time to retrieve top 20 per query	1 sec.
System RAM	4 GB
Size of on-disk structures	54 GB
Approx system cost	$28200

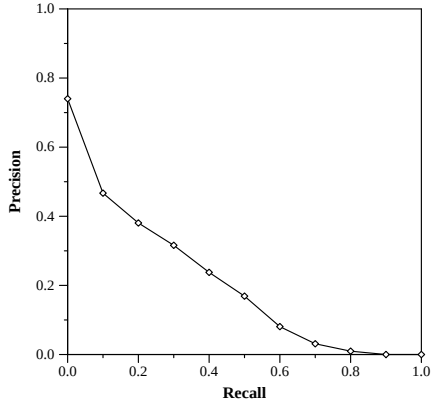
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	49000
Relevant:	10617
Rel-ret:	5011

Recall Level Averages	
Recall	Precision
0.00	0.7401
0.10	0.4669
0.20	0.3806
0.30	0.3160
0.40	0.2377
0.50	0.1689
0.60	0.0807
0.70	0.0311
0.80	0.0097
0.90	0.0000
1.00	0.0000

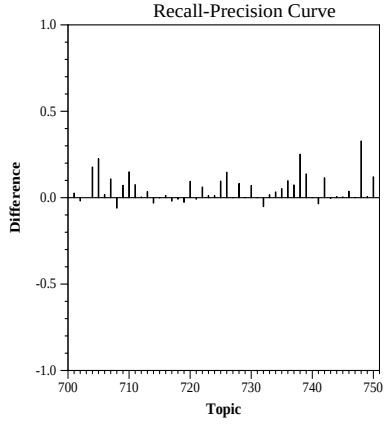
Mean average precision	
non-interpolated	0.1912

Document Level Averages	
	Precision
At 5 docs	0.4694
At 10 docs	0.4571
At 15 docs	0.4395
At 20 docs	0.4255
At 30 docs	0.4170
At 100 docs	0.3533
At 200 docs	0.2756
At 500 docs	0.1650
At 1000 docs	0.1023

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.2719



Recall-Precision Curve



Difference from Median Average Precision per Topic

Terabyte Track results — Microsoft Research Asia

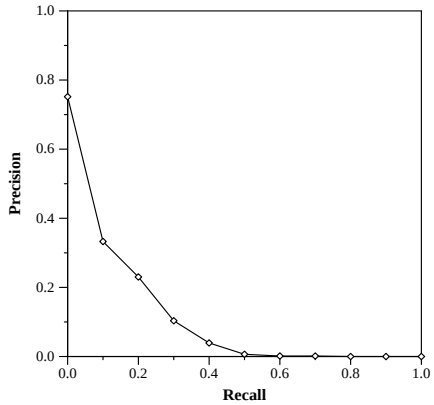
Summary Statistics	
Run ID	MSRAt2
Run Type	automatic
Fields Used	title
Run Description	NoLinksUsed-NoAnchorTextUsed-DocStructUsed
Time to index	700 min.
Time to retrieve top 20 per query	1 sec.
System RAM	4 GB
Size of on-disk structures	54 GB
Approx system cost	$28200
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	49000
Relevant:	10617
Rel-ret:	3011

Recall Level Averages	
Recall	Precision
0.00	0.7517
0.10	0.3328
0.20	0.2302
0.30	0.1035
0.40	0.0392
0.50	0.0065
0.60	0.0014
0.70	0.0014
0.80	0.0000
0.90	0.0000
1.00	0.0000

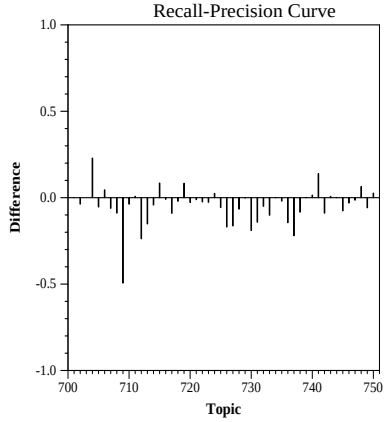
Mean average precision	
non-interpolated	0.0920

Document Level Averages	
	Precision
At 5 docs	0.4122
At 10 docs	0.3469
At 15 docs	0.3279
At 20 docs	0.3082
At 30 docs	0.3007
At 100 docs	0.2612
At 200 docs	0.1839
At 500 docs	0.0978
At 1000 docs	0.0614

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.1604



Recall-Precision Curve



Difference from Median Average Precision per Topic

Terabyte Track results — Microsoft Research Asia

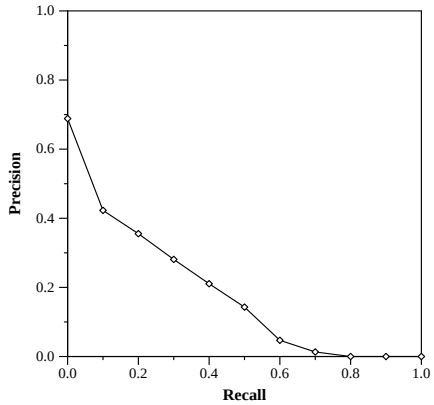
Summary Statistics	
Run ID	MSRAt3
Run Type	automatic
Fields Used	title
Run Description	NoLinksUsed-AnchorTextUsed-DocStructUsed
Time to index	700 min.
Time to retrieve top 20 per query	1 sec.
System RAM	4 GB
Size of on-disk structures	54 GB
Approx system cost	$28200
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	49000
Relevant:	10617
Rel-ret:	4797

Recall Level Averages	
Recall	Precision
0.00	0.6884
0.10	0.4228
0.20	0.3554
0.30	0.2810
0.40	0.2106
0.50	0.1430
0.60	0.0469
0.70	0.0132
0.80	0.0000
0.90	0.0000
1.00	0.0000

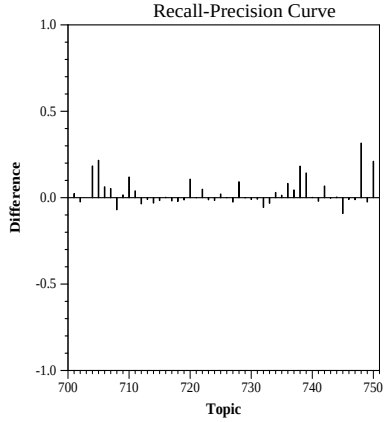
Mean average precision	
non-interpolated	0.1712

Document Level Averages	
	Precision
At 5 docs	0.4490
At 10 docs	0.4327
At 15 docs	0.4231
At 20 docs	0.4000
At 30 docs	0.3891
At 100 docs	0.3312
At 200 docs	0.2610
At 500 docs	0.1586
At 1000 docs	0.0979

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.2525



Recall-Precision Curve



Difference from Median Average Precision per Topic

Terabyte Track results — Microsoft Research Asia

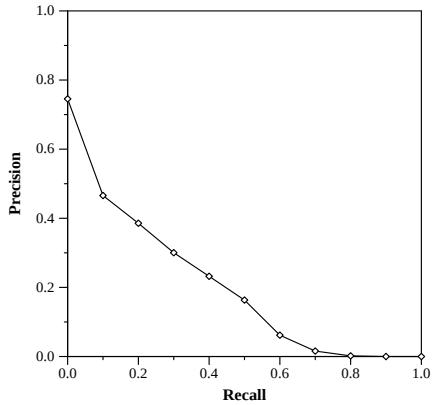
Summary Statistics	
Run ID	MSRAt4
Run Type	automatic
Fields Used	title
Run Description	NoLinksUsed-AnchorTextUsed-DocStructUsed
Time to index	700 min.
Time to retrieve top 20 per query	1 sec.
System RAM	4 GB
Size of on-disk structures	54 GB
Approx system cost	$28200
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	49000
Relevant:	10617
Rel-ret:	4932

Recall Level Averages	
Recall	Precision
0.00	0.7455
0.10	0.4658
0.20	0.3856
0.30	0.3004
0.40	0.2321
0.50	0.1633
0.60	0.0617
0.70	0.0158
0.80	0.0019
0.90	0.0000
1.00	0.0000

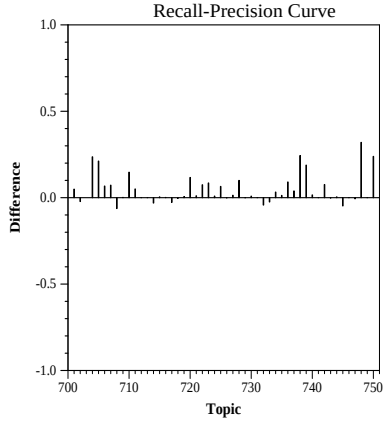
Mean average precision	
non-interpolated	0.1877

Document Level Averages	
	Precision
At 5 docs	0.4898
At 10 docs	0.4959
At 15 docs	0.4503
At 20 docs	0.4337
At 30 docs	0.4122
At 100 docs	0.3469
At 200 docs	0.2741
At 500 docs	0.1661
At 1000 docs	0.1007

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.2741



Recall-Precision Curve



Difference from Median Average Precision per Topic

Terabyte Track results — Microsoft Research Asia

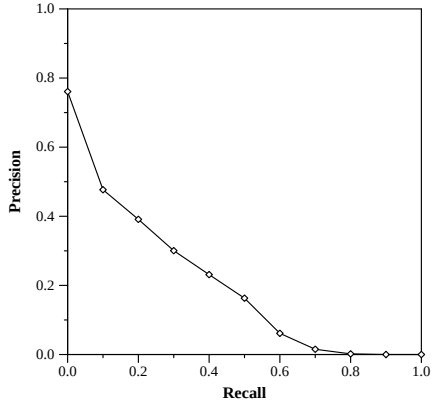
Summary Statistics	
Run ID	MSRAt5
Run Type	automatic
Fields Used	title
Run Description	NoLinksUsed-AnchorTextUsed-DocStructUsed
Time to index	700 min.
Time to retrieve top 20 per query	1 sec.
System RAM	4 GB
Size of on-disk structures	54 GB
Approx system cost	$28200
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	49000
Relevant:	10617
Rel-ret:	4933

Recall Level Averages	
Recall	Precision
0.00	0.7608
0.10	0.4768
0.20	0.3911
0.30	0.3005
0.40	0.2312
0.50	0.1630
0.60	0.0613
0.70	0.0152
0.80	0.0019
0.90	0.0000
1.00	0.0000

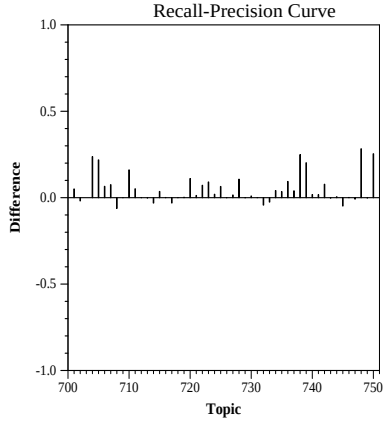
Mean average precision	
non-interpolated	0.1902

Document Level Averages	
	Precision
At 5 docs	0.5102
At 10 docs	0.4816
At 15 docs	0.4558
At 20 docs	0.4418
At 30 docs	0.4156
At 100 docs	0.3488
At 200 docs	0.2750
At 500 docs	0.1660
At 1000 docs	0.1007

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.2749



Recall-Precision Curve



Difference from Median Average Precision per Topic