Terabyte Track results — Max-Planck-Institute for Computer Science

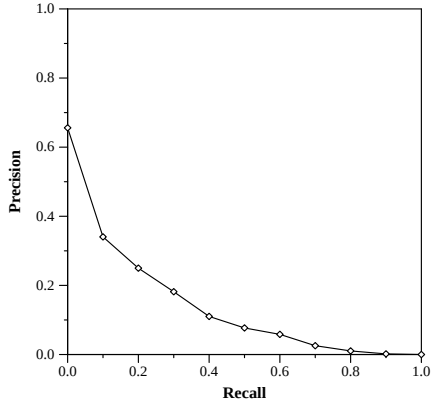
Summary Statistics	
Run ID	mpi04tb07
Run Type	automatic
Fields Used	title
Run Description	LinksUsed-NoAnchorTextUsed-DocStructUsed
Time to index	2520 min.
Time to retrieve top 20 per query	6 sec.
System RAM	2 GB
Size of on-disk structures	479 GB
Approx system cost	$15000
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	48266
Relevant:	10617
Rel-ret:	4143

Recall Level Averages	
Recall	Precision
0.00	0.6559
0.10	0.3404
0.20	0.2499
0.30	0.1817
0.40	0.1102
0.50	0.0769
0.60	0.0583
0.70	0.0255
0.80	0.0104
0.90	0.0018
1.00	0.0000

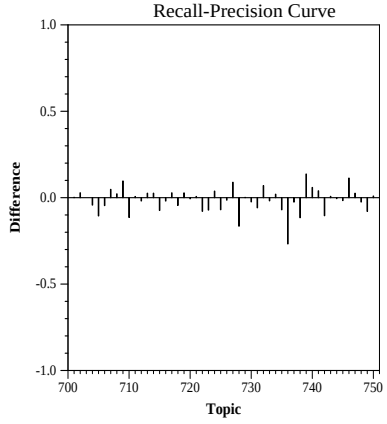
Mean average precision	
non-interpolated	0.1249

Document Level Averages	
	Precision
At 5 docs	0.4245
At 10 docs	0.4122
At 15 docs	0.3864
At 20 docs	0.3663
At 30 docs	0.3388
At 100 docs	0.2667
At 200 docs	0.2033
At 500 docs	0.1306
At 1000 docs	0.0846

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.2108



Recall-Precision Curve



Difference from Median Average Precision per Topic

Terabyte Track results — Max-Planck-Institute for Computer Science

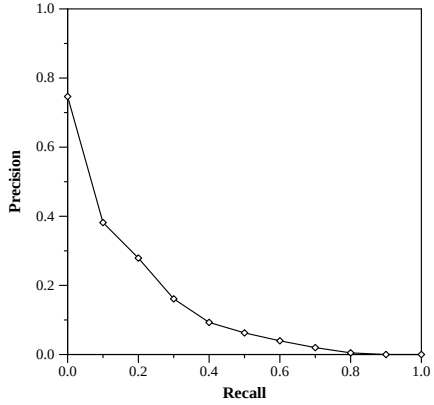
Summary Statistics	
Run ID	mpi04tb09
Run Type	automatic
Fields Used	title, description
Run Description	LinksUsed-NoAnchorTextUsed-DocStructUsed
Time to index	2520 min.
Time to retrieve top 20 per query	9 sec.
System RAM	2 GB
Size of on-disk structures	479 GB
Approx system cost	$15000
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	48120
Relevant:	10617
Rel-ret:	3908

Recall Level Averages	
Recall	Precision
0.00	0.7465
0.10	0.3818
0.20	0.2791
0.30	0.1610
0.40	0.0929
0.50	0.0626
0.60	0.0397
0.70	0.0200
0.80	0.0048
0.90	0.0000
1.00	0.0000

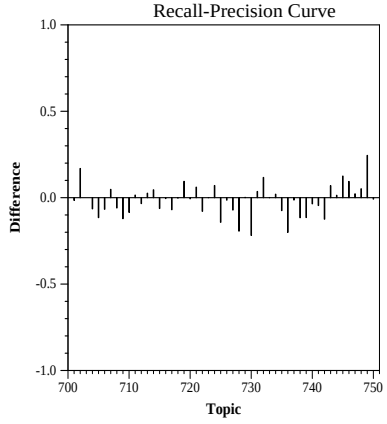
Mean average precision	
non-interpolated	0.1232

Document Level Averages	
	Precision
At 5 docs	0.4367
At 10 docs	0.4082
At 15 docs	0.3755
At 20 docs	0.3653
At 30 docs	0.3415
At 100 docs	0.2927
At 200 docs	0.2049
At 500 docs	0.1241
At 1000 docs	0.0798

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.2043



Recall-Precision Curve



Difference from Median Average Precision per Topic

Terabyte Track results — Max-Planck-Institute for Computer Science

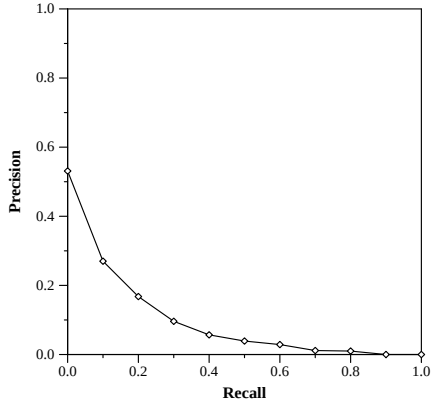
Summary Statistics	
Run ID	mpi04tb101
Run Type	automatic
Fields Used	title, description
Run Description	LinksUsed-NoAnchorTextUsed-NoDocStructUsed
Time to index	2520 min.
Time to retrieve top 20 per query	9 sec.
System RAM	6 GB
Size of on-disk structures	479 GB
Approx system cost	$15000
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	49000
Relevant:	10617
Rel-ret:	2896

Recall Level Averages	
Recall	Precision
0.00	0.5310
0.10	0.2700
0.20	0.1677
0.30	0.0961
0.40	0.0569
0.50	0.0390
0.60	0.0288
0.70	0.0115
0.80	0.0100
0.90	0.0000
1.00	0.0000

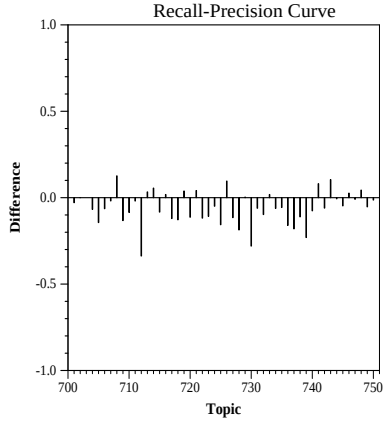
Mean average precision	
non-interpolated	0.0811

Document Level Averages	
	Precision
At 5 docs	0.3061
At 10 docs	0.2878
At 15 docs	0.2830
At 20 docs	0.2704
At 30 docs	0.2537
At 100 docs	0.1902
At 200 docs	0.1433
At 500 docs	0.0913
At 1000 docs	0.0591

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.1453



Recall-Precision Curve



Difference from Median Average Precision per Topic

Terabyte Track results — Max-Planck-Institute for Computer Science

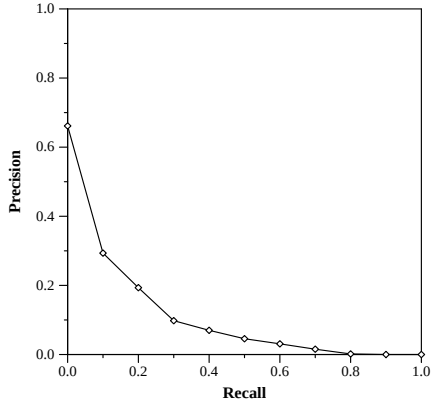
Summary Statistics	
Run ID	mpi04tb81
Run Type	automatic
Fields Used	title, description
Run Description	LinksUsed-NoAnchorTextUsed-NoDocStructUsed
Time to index	2520 min.
Time to retrieve top 20 per query	9 sec.
System RAM	6 GB
Size of on-disk structures	479 GB
Approx system cost	$15000
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	49000
Relevant:	10617
Rel-ret:	3076

Recall Level Averages	
Recall	Precision
0.00	0.6614
0.10	0.2934
0.20	0.1934
0.30	0.0979
0.40	0.0700
0.50	0.0457
0.60	0.0309
0.70	0.0154
0.80	0.0017
0.90	0.0000
1.00	0.0000

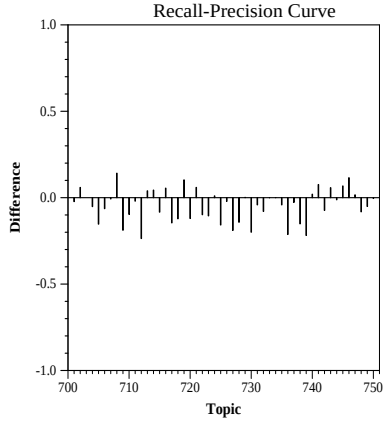
Mean average precision	
non-interpolated	0.0922

Document Level Averages	
	Precision
At 5 docs	0.4082
At 10 docs	0.3653
At 15 docs	0.3483
At 20 docs	0.3367
At 30 docs	0.3150
At 100 docs	0.2045
At 200 docs	0.1485
At 500 docs	0.0954
At 1000 docs	0.0628

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.1540



Recall-Precision Curve



Difference from Median Average Precision per Topic

Terabyte Track results — Max-Planck-Institute for Computer Science

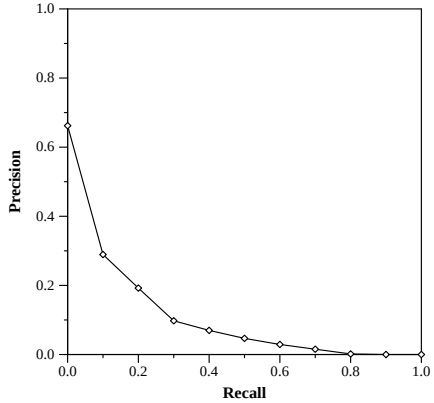
Summary Statistics	
Run ID	mpi04tb91
Run Type	automatic
Fields Used	title, description
Run Description	LinksUsed-NoAnchorTextUsed-NoDocStructUsed
Time to index	2520 min.
Time to retrieve top 20 per query	9 sec.
System RAM	6 GB
Size of on-disk structures	479 GB
Approx system cost	$15000
Number of Topics:	49
	Total number of documents over all topics
Retrieved:	49000
Relevant:	10617
Rel-ret:	3071

Recall Level Averages	
Recall	Precision
0.00	0.6621
0.10	0.2893
0.20	0.1923
0.30	0.0976
0.40	0.0698
0.50	0.0467
0.60	0.0291
0.70	0.0154
0.80	0.0017
0.90	0.0000
1.00	0.0000

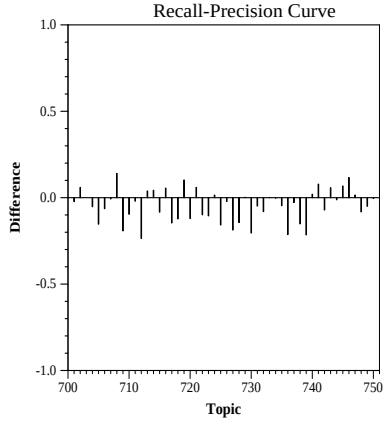
Mean average precision	
non-interpolated	0.0919

Document Level Averages	
	Precision
At 5 docs	0.4122
At 10 docs	0.3612
At 15 docs	0.3510
At 20 docs	0.3357
At 30 docs	0.3129
At 100 docs	0.2051
At 200 docs	0.1485
At 500 docs	0.0949
At 1000 docs	0.0627

R-Precision: precision after R (number relevant) documents retrieved	
Exact	0.1535



Recall-Precision Curve



Difference from Median Average Precision per Topic