TREC_EVAL EVALUATION REPORT

Ranked Retrieval Evaluation

Terms used for analysis:

A. Recall

Recall is a measure of the ability of a system to present all relevant items.

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

B. Precision at 11 standard recall levels

It is the input for plotting recall-precision graph. These values are used to compare the performance of different systems.

C. Average precision over all relevant documents, interpolated (ircl_prn)

ircl prn are the interpolation precisions at different values of recall; they can be used to draw a Precision-Recall graph.

D. Precision at 9 document cutoff values (k-Precision)

The precision computed after a given number of documents have been retrieved reflects the actual measured system performance as a user might see it. Each document precision average is computed by summing the precisions at the specified document cutoff value and dividing by the number of topics (150).

E. R-Precision

R-Precision is the precision after R documents have been retrieved, where R is the number of relevant documents for the topic. It de-emphasizes the exact ranking of the retrieved relevant documents. The average R-Precision for a run is computed by taking the mean of the R-Precisions of the individual topics in the run.

F. F measure

F-measure (or F1 score) is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0.

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

The general formula for positive real β is:

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}.$$

$$E = 1 - \left(\frac{\alpha}{P} + \frac{1-\alpha}{R}\right)^{-1}.$$

Their relationship is $F_\beta = 1 - E$ where $\alpha = \frac{1}{1+\beta^2}$.

G. MAP

The Mean Average Precision (MAP) is the arithmetic mean of the average precision values for an information retrieval system over a set of n query topics. It can be expressed as follows:

$$MAP = \frac{1}{n} \sum_{n} AP_n$$

H. GMAP

The Geometric Mean Average Precision (GMAP) is the geometric mean of the average precision values for an information retrieval system over a set of n query topics. It is expressed as follows:

$$\text{GMAP} = \sqrt[n]{\prod_{n} AP_n}$$

where AP represents the Average Precision value for a given topic from the evaluation set of n topics. An alternate calculation method expresses GMAP as an arithmetic mean of logs:

$$\text{GMAP} = \exp \frac{1}{n} \sum_{n} \log AP_n$$

I. Bpref

bpref computes a preference relation of whether judged relevant documents are retrieved ahead of judged irrelevant documents. Thus, it is based on the relative ranks of judged documents only. The bpref measure is defined as

$$\text{bpref} = \frac{1}{R} \sum_{r} (1 - \frac{|n \text{ ranked higher than } r|}{\min(R, N)})$$

J. Reciprocal rank

Reciprocal Rank (RR) is the reciprocal of the first relevant document's rank in the ranked list returned for a topic. E.g. the first relevant document is ranked as No. 4, RR = ¼ = 0.25.

PART 1

For the trec_eval program, the command that we have used consists of the two sets of input that are:

trec eval [-q] [-a] trec_qrel_file trec_results_file
trec_eval.9.0/trec_eval test-data/qrels.trec6-8.nocr result/result.out

1) Ranked list of documents generated for tf-idf and bm25 modules by BatchSearch.java (tf-idf_result.out and result.out) (Marked by green)

   The format of these files is:

query-number	Q0	document-id	rank	score	Exp
448	Q0	FBIS3-42486	463	0.24165525	default

   query-number - the number of the query,
   document-id - the external ID for the retrieved document,
   score - the score that the retrieval system creates for that document against that query.
   Q0 (Q zero) and Exp - constants that are used by some evaluation software

2) the qrel file that contains for each query the set of all documents judged as relevant or non-relevant. (qrels.trec6-8.nocr)

query-number	0	document-id	relevance
301	0	FBIS3-20360	1

   query-number - the number of the query,
   document-id - the external ID for the judged documents, 0 is a constant and
   relevance - the relevance assigned to the document for the particular query; relevance is either 0 (non- relevant) or 1 (relevant)

The results obtained by running trec_eval are

num_ret Total number of documents retrieved over all queries
num_rel Total number of relevant documents over all queries
num_rel_ret Total number of relevant documents retrieved over all queries
map Mean Average Precision (MAP)
gm_ap Average Precision. Geometric Mean, q_score=log(MAX(map,.00001))
R-prec R-Precision (Precision after R (= num-rel for topic) documents retrieved)
bpref Binary Preference, top R judged nonrel
recip_rank Reciprical rank of top relevant document
ircl_prn.0.00 Interpolated Recall - Precision Averages at 0.00 recall
ircl_prn.0.10 Interpolated Recall - Precision Averages at 0.10 recall
ircl_prn.0.20 Interpolated Recall - Precision Averages at 0.20 recall
ircl_prn.0.30 Interpolated Recall - Precision Averages at 0.30 recall
ircl_prn.0.40 Interpolated Recall - Precision Averages at 0.40 recall
ircl_prn.0.50 Interpolated Recall - Precision Averages at 0.50 recall
ircl_prn.0.60 Interpolated Recall - Precision Averages at 0.60 recall
ircl_prn.0.70 Interpolated Recall - Precision Averages at 0.70 recall
ircl_prn.0.80 Interpolated Recall - Precision Averages at 0.80 recall
ircl_prn.0.90 Interpolated Recall - Precision Averages at 0.90 recall
ircl_prn.1.00 Interpolated Recall - Precision Averages at 1.00 recall
P5 Precision after 5 docs retrieved
P10 Precision after 10 docs retrieved
P15 Precision after 15 docs retrieved
P20 Precision after 20 docs retrieved
P30 Precision after 30 docs retrieved

P100 Precision after 100 docs retrieved
P200 Precision after 200 docs retrieved
P500 Precision after 500 docs retrieved
P1000 Precision after 1000 docs retrieved

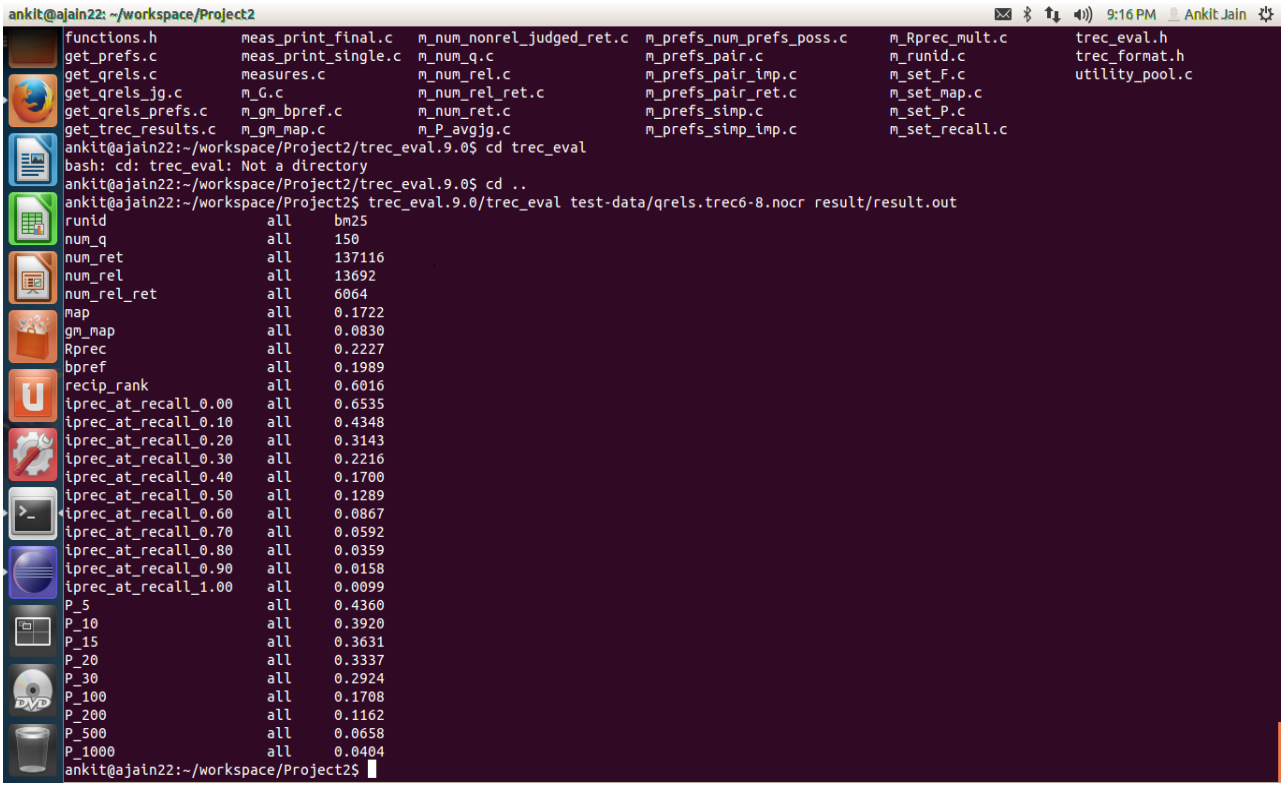
1000 documents being retrieved for each query. Therefore, summing up the results of these queries:

ANALYSIS OF RESULTS FOR TITLE QUERY

1) BM25 Model:

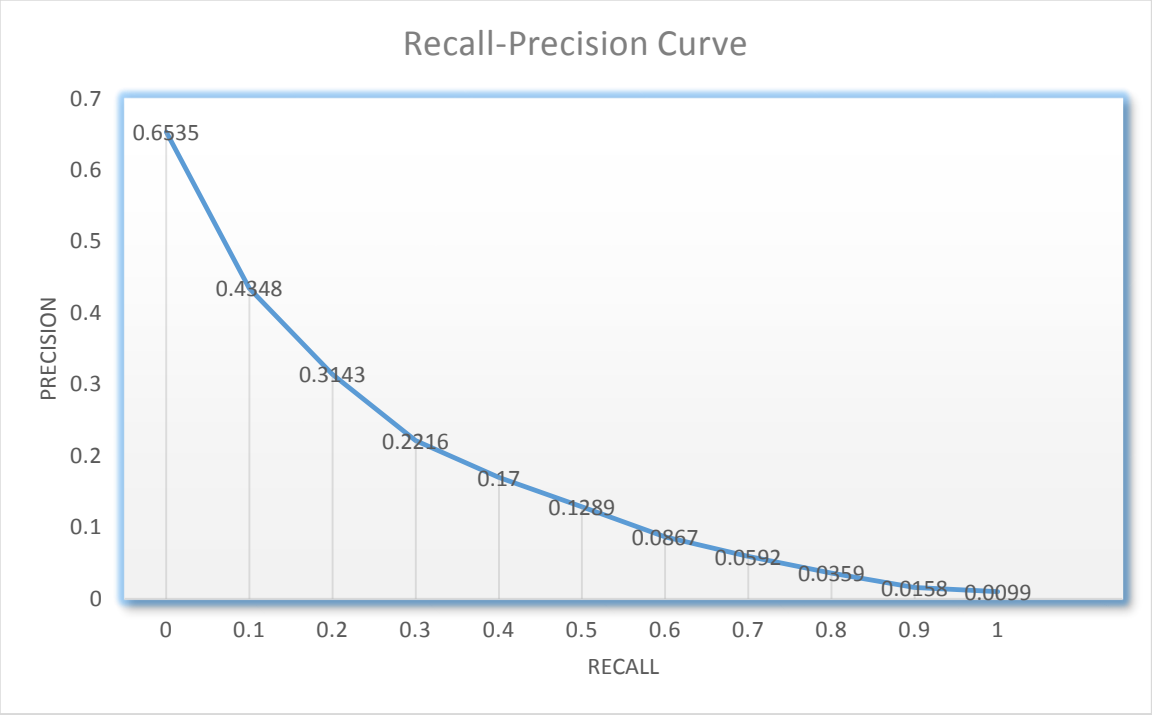
Query Output file:

result.out



Summary Statistics Table

Number of topics (num_q)	150
Total number of documents retrieved over all queries (num_ret)	137116
Total number of relevant documents over all queries (num_rel)	13692
Total number of relevant documents retrieved over all queries (num_rel_ret)	6064

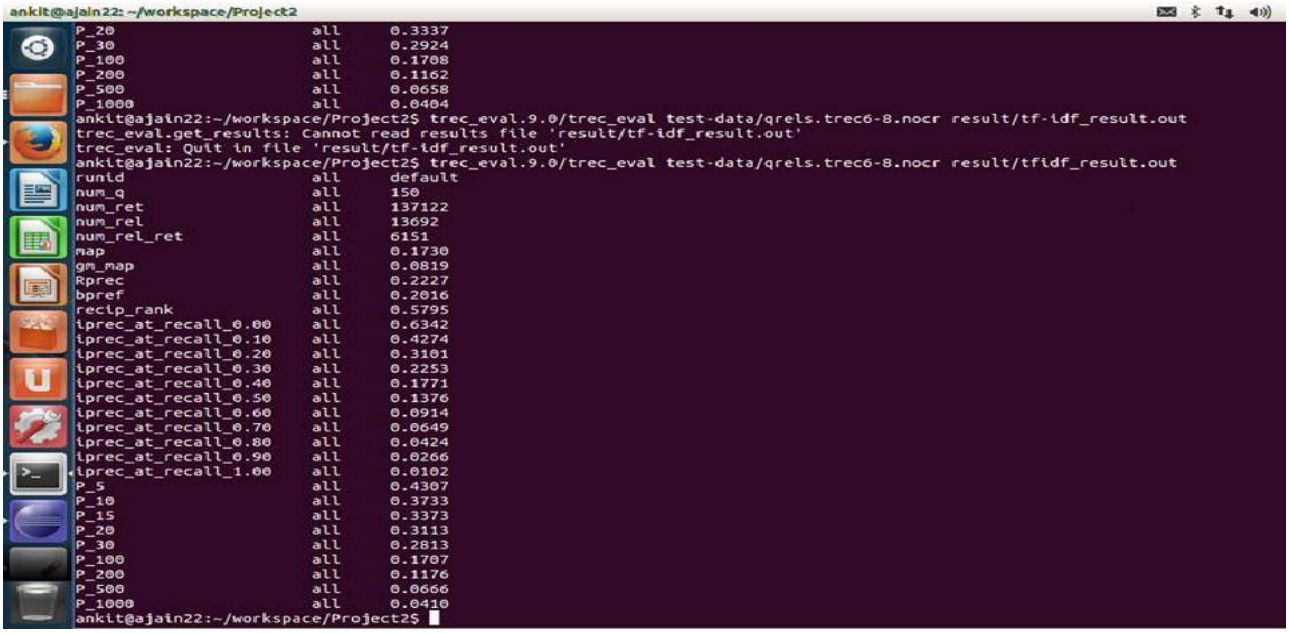


Analysis: The graph slopes downward from left to right which means as relevant documents are retrieved (recall increases), the more non relevant documents are retrieved (precision decreases).

2) TF-IDF (Vector Space) Model:

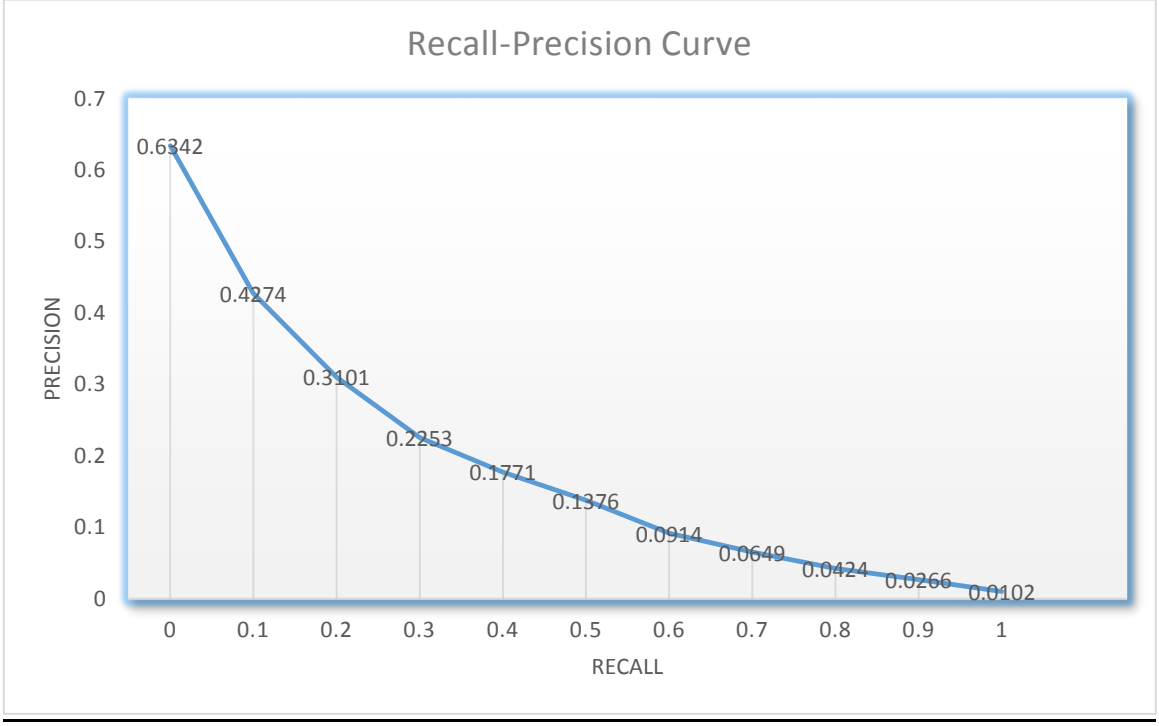
Query Output file:

tfidf_result.out



Summary Statistics Table

Number of topics (num_q)	150
Total number of documents retrieved over all queries (num_ret)	137122
Total number of relevant documents over all queries (num_rel)	13692
Total number of relevant documents retrieved over all queries (num_rel_ret)	6151



Analysis: The graph slopes downward from left to right which means as relevant documents are retrieved (recall increases), the more non relevant documents are retrieved (precision decreases).

COMPARISON BETWEEN VECTOR SPACE MODEL AND BM25 MODEL

Recall Level Precision Averages Comparison

Recall Level Precision Averages		
Recall	**Precision of BM25**	**Precision of TF-IDF**
0	0.6535	0.6342
0.1	0.4348	0.4274
0.2	0.3143	0.3101
0.3	0.2216	0.2253
0.4	0.17	0.1771
0.5	0.1289	0.1376
0.6	0.0867	0.0914
0.7	0.0592	0.0649
0.8	0.0359	0.0424
0.9	0.0158	0.0266
1	0.0099	0.0102

Average Precision over all Relevant Docs		
Interpolated (ircl_prn)	0.1936	0.1952

Document Level Precision Averages Comparison

Document Level Precision Averages		
No. of Docs (R)	**Precision of BM25**	**Precision of TF-IDF**
At 5 docs	0.4307	0.4307
At 10 docs	0.3733	0.3733
At 15 docs	0.3373	0.3373
At 20 docs	0.3113	0.3113
At 30 docs	0.2813	0.2813
At 100 docs	0.1707	0.1707
At 200 docs	0.1176	0.1176
At 500 docs	0.0666	0.0666
At 1000 docs	0.041	0.041

K-Precision (Precision after R docs Retrieved)		
Exact	0.2366	0.2366

Comparison on the basis of different measures

	BM25	VSM
Interpolated Average Precision	0.1936	0.1952
Precision	0.0442	0.0448
Recall	0.4428	0.4492
F measure	0.0802	0.0813
k-precision	0.2456	0.2366
R-precision	0.2227	0.2227
MAP	0.1722	0.173
GMAP	0.083	0.0819
bpref	0.1989	0.2016
Reciprocal Rank	0.60	0.57

We have used the following evaluation parameters:

1. **Interpolated Average Precision:** The interpolated precision at the j-th standard recall level is the maximum known precision at any recall level between the j-th and (j + 1)-th level:

$$P(r_j) = \max_{r_j \le r \le r_{j+1}} P(r)$$

2. **Precision:** Precision is the number of relevant documents in the ranked list returned for a topic. Naturally, the model which gives higher Precision is evaluated as better. From the analysis we see that VSM is better.

3. **Recall:** Recall refers to the fraction of relevant documents which have been retrieved. So, on the basis of Recall we can say that the model showing higher value of Recall can be considered as better. Therefore, VSM is better.

   ***Note:** There is a tradeoff between recall and precision. One can increase recall by retrieving more documents but this can decrease precision.

4. **F measure:** For, High alpha: Precision is more important. Low alpha: Recall is more important Maximum value of F0:5-measure (or F-measure for short) is a good indication of best P/R compromise. F-measure is an approximation of cross-over point of precision and recall. In our case, F Measure of VSM model is greater which signifies greater precision value or smaller recall value. Both these conditions favor a good IR retrieval model. Therefore, VSM is better.

5. **k-Precision:** Precision at k documents is still a useful metric but fails to take into account the positions of the relevant documents among the top k.
   User is mainly concerned with the precision of early documents received.

k-Precision	At 5 docs	At 10 docs	At 15 docs	At 20 docs	At 30 docs	At 100 docs
BM25	0.436	0.392	0.3631	0.3337	0.2924	0.1708
VSM	0.4307	0.3733	0.3373	0.3113	0.2813	0.1707

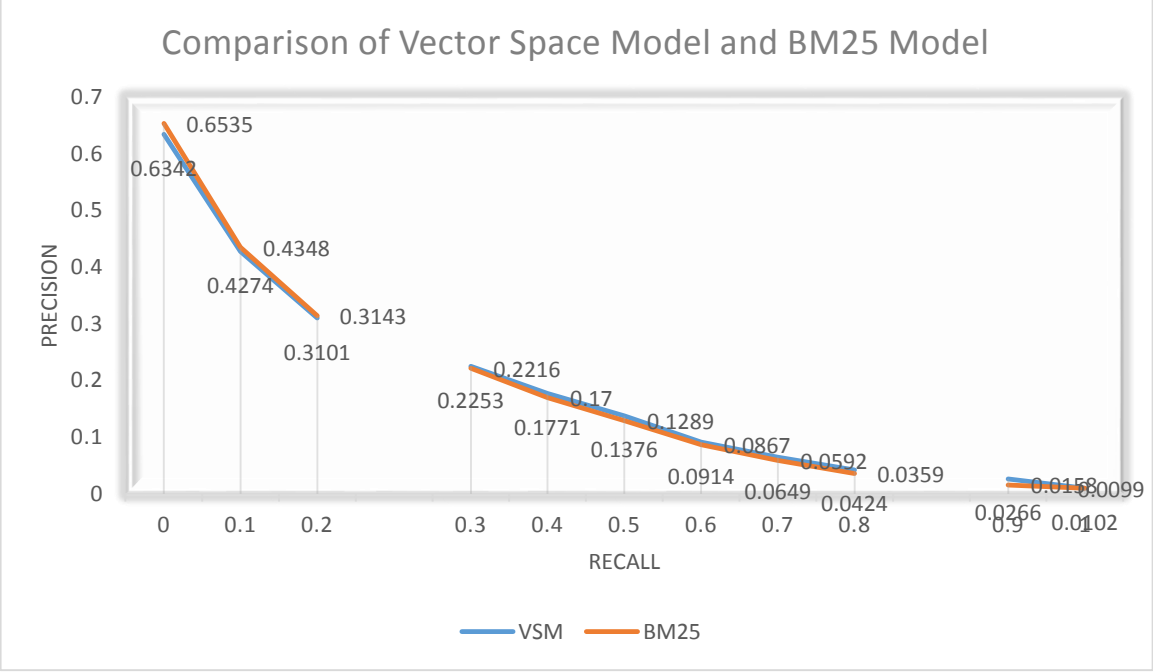
   It is clear that Precision for top 100 results is marginally greater in case of BM25 model, i.e more relevant documents are retrieved in case of Okapi for top 100 documents. Therefore, Okapi is better.

6. **R-Precision:** Precision at the R-th position in the ranking of results for a query that has R relevant documents. The model for which R precision is greater indicates that more number of relevant documents have been retrieved for that model till the R-th position.

7. **MAP:** MAP favors systems which return relevant documents fast i.e. Precision-biased. Higher is its value, better is the model. In our case, MAP for BM25 is 0.1722 and MAP for VSM is 0.1730.
   Therefore, VSM is better.

8. **Reciprocal Rank:** RR captures how early we get relevant result in ranking reciprocal rank of ranked results of a query.
   perfect = 1 → worse → 0
   In our case, for BM25 RR = 0.60 and for VSM, RR = 0.57
   i.e. in Okapi BM25 model, we get more relevant results (score is greater for top k documents) as compared to VSM. Therefore, Okapi is better.

COMPARISON ON THE BASIS OF PRECISION RECALL CURVE

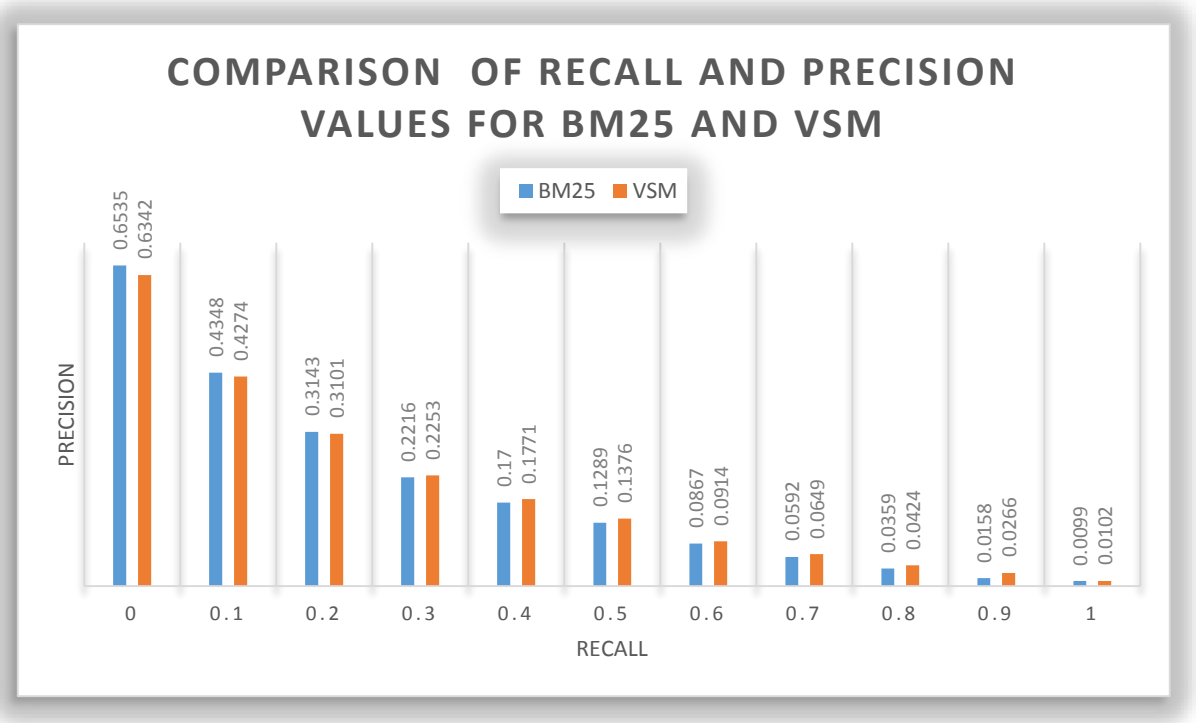


The plots of different runs can be superimposed on the same graph to determine which run is superior. Curves closest to the upper right-hand corner of the graph (where recall and precision are maximized) indicate the best performance.

Comparisons are best made in three different recall ranges:

0 to 0.2 --- high precision (BM25 is better)
0.2 to 0.8 -- middle recall (VSM is better)
0.8 to 1 -- high recall performance (VSM is better)

Q.1) Which model had best early precision?
Best Early Precision: BM25 (Okapi)



SUMMARY OF COMPARISONS

Parameter	Model which performed better
Precision	VSM
Recall	VSM
F measure	VSM
k-precision	BM25
MAP	VSM
Reciprocal Rank	BM25
Precision Recall Curve	VSM (For Recall) BM25(For Precision)

Which performed best overall?

However, it is difficult to judge, which model performed best, but based on the information need of the used, decision can be made. In this case, the object is to check if the information that the user is looking for is present in the collection or not. Query consist of titles of the documents. User would like to retrieve the needed document with best precision and in the top 10 search results.
Here, **high early precision is better for BM25**, while the overall precision is better for VSM. But, the user would be interested in the top results, for which BM25 has got better precision.

So, it can be inferred that BM25 performs good initially, but based on other parameters like F-Measure, Precision, Recall, MAP and Precision recall curve, **the overall performance of VSM is better.**

Which queries had particularly poor performance and why?

From the qrel file we can find out the relevancy of the document for the query. So, if we can find out the relevant documents for a query and out of those queries, we can find out the minimum score, then we can get the poorest performing query (Because the contribution of this query in the score would be minimum).

In particular, **query #313** – "Magnetic Levitation Maglev" gives the most poor performance because it has the **least score** of 1.7710232 for BM25 model and 0.020710444 for VSM (default) model.

PART 2

For Part 2, we have used the following **Query Expansion** approaches for IR system evaluation:

1) We have written a Java code to create narrative and descriptive queries from topics.301-450 documents. These query files are then used as an argument for BatchSearch.java file for generating the result file for both the models. This result file is then used with trec_eval command to compare it with q_rel file and get the values for different evaluation criterion.

2) In the second approach, we are using the query file which picks up the terms from title, description and narrative sections of topics.301-450 file. This approach focusses on making the query more powerful to improve the overall performance of the IR model.

At the end we will check the evaluation parameters for both these approaches and come out with a conclusion for the best approach**.**

APPROACH 1

ALGORITHM FOR BASIC QUERY PROCESSING

1. Read the topics file
2. For each topic
3. Extract the query number
4. Fetch title/narrative/description corresponding to that query number
5. Remove punctuations, and other non-alpha numeric characters from the text
6. Remove the stop words from this processed text
7. Write the resultant query into separate query files
8. Use this query file for generating ranks and scores of documents

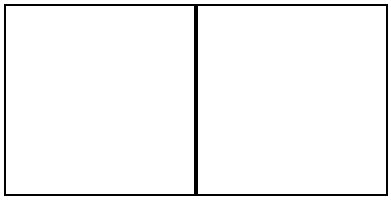
QUERY PROCESSING CODE USING JAVA


QueryExtract.java


NarrativeQuery.java

ACTUAL RESULTS

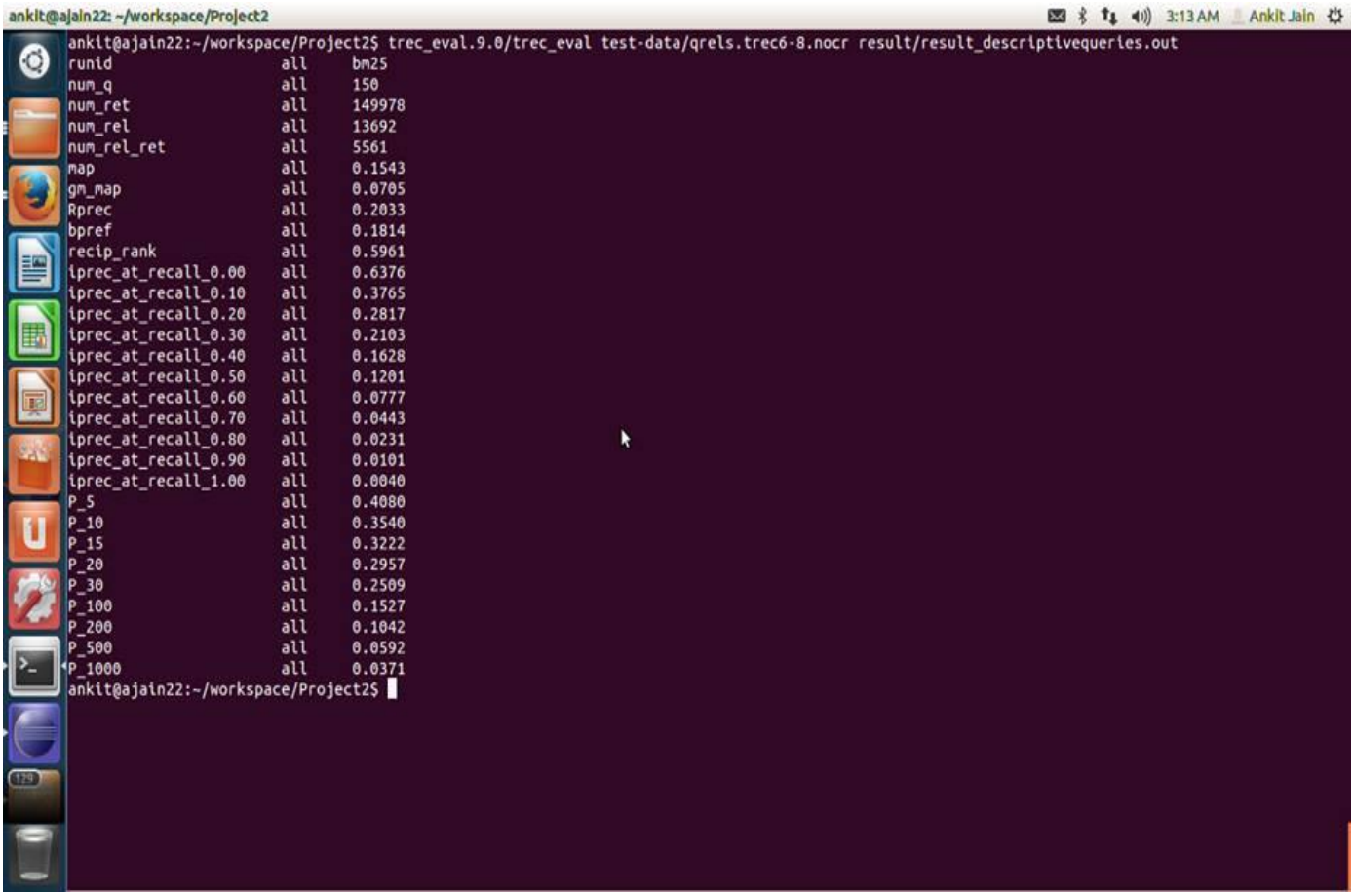
i) Narrative and Descriptive Query Files generated from the above code are:



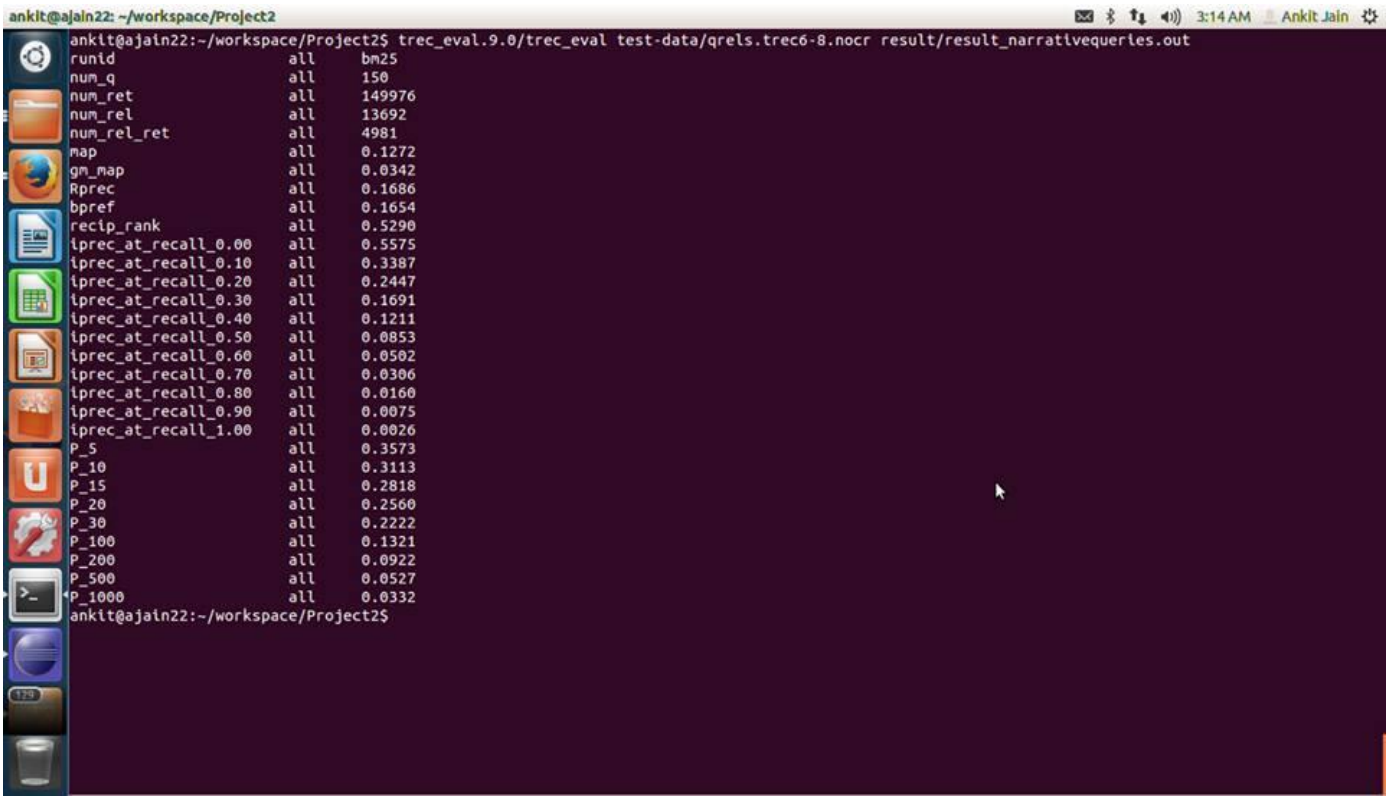
ii) Running BatchSearch.java to extract result for BM25 for descriptive and narrative queries:

iii) Running BatchSearch.java to extract result for VSM for descriptive and narrative queries:

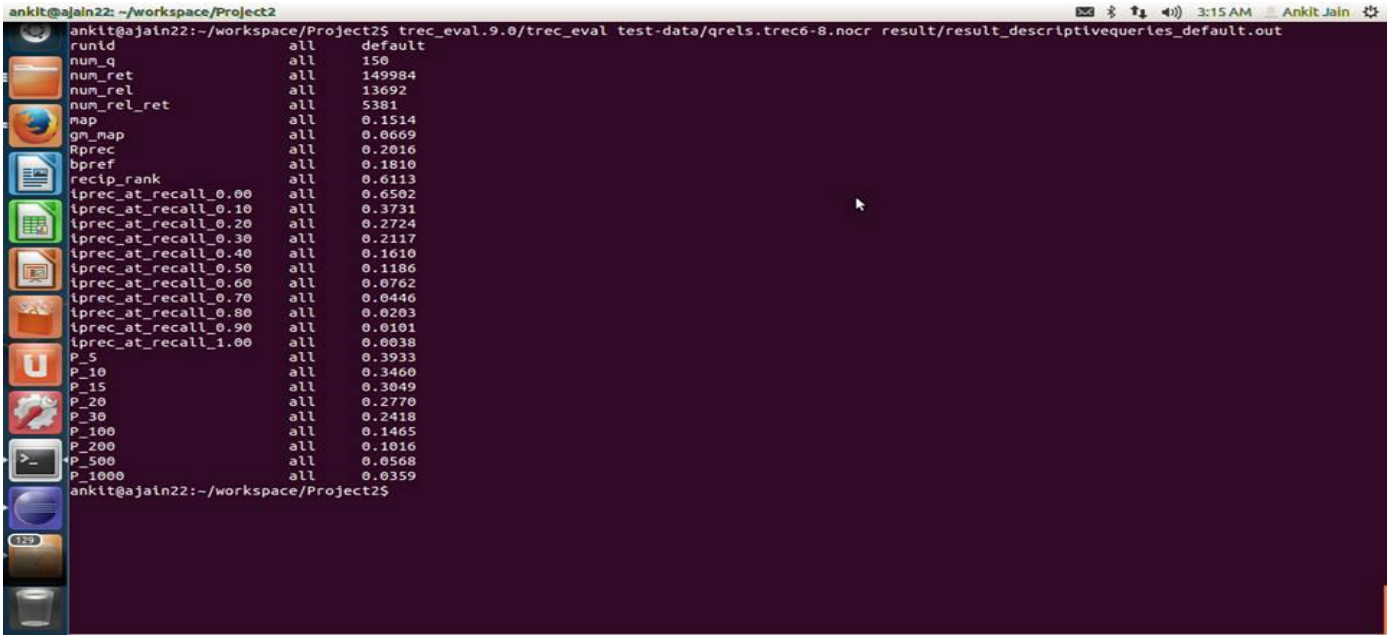
iv) Running the below command for BM25 for descriptive queries



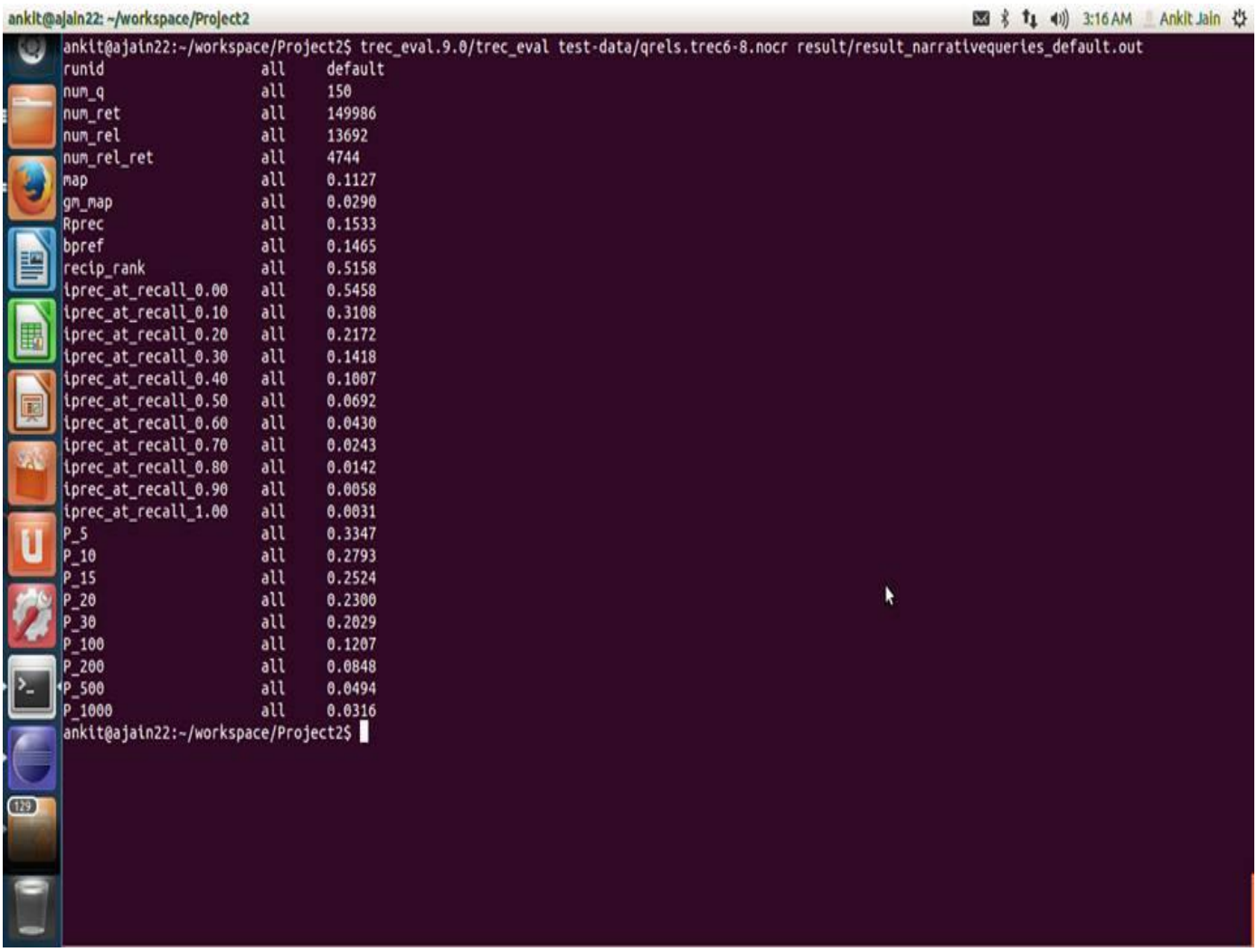
v) Running the command for BM25 for narrative queries



vi) Running the command for VSM for descriptive queries



vii) Running the command for VSM for narrative queries

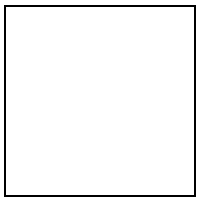

ankit@ajain22: ~/workspace/Project2
3:16 AM
Ankit Jain
ankit@ajain22:~/workspace/Project2$ trec_eval.9.0/trec_eval test-data/qrels.trec6-8.nocr result/result_narrativequeries_default.out
runid all default
num_q all 150
num_ret all 149986
num_rel all 13692
num_rel_ret all 4744
map all 0.1127
gm_map all 0.0290
Rprec all 0.1533
bpref all 0.1465
recip_rank all 0.5158
iprec_at_recall_0.00 all 0.5458
iprec_at_recall_0.10 all 0.3108
iprec_at_recall_0.20 all 0.2172
iprec_at_recall_0.30 all 0.1418
iprec_at_recall_0.40 all 0.1007
iprec_at_recall_0.50 all 0.0692
iprec_at_recall_0.60 all 0.0430
iprec_at_recall_0.70 all 0.0243
iprec_at_recall_0.80 all 0.0142
iprec_at_recall_0.90 all 0.0058
iprec_at_recall_1.00 all 0.0031
P_5 all 0.3347
P_10 all 0.2793
P_15 all 0.2524
P_20 all 0.2300
P_30 all 0.2029
P_100 all 0.1207
P_200 all 0.0848
P_500 all 0.0494
P_1000 all 0.0316
ankit@ajain22:~/workspace/Project2$

APPROACH 2

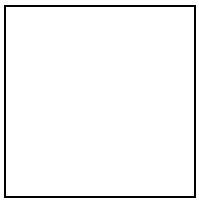
QUERY PROCESSING CODE USING JAVA

ACTUAL RESULTS

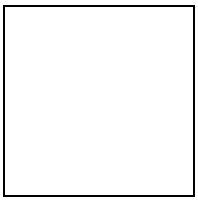
i) Query File generated from the above code are:



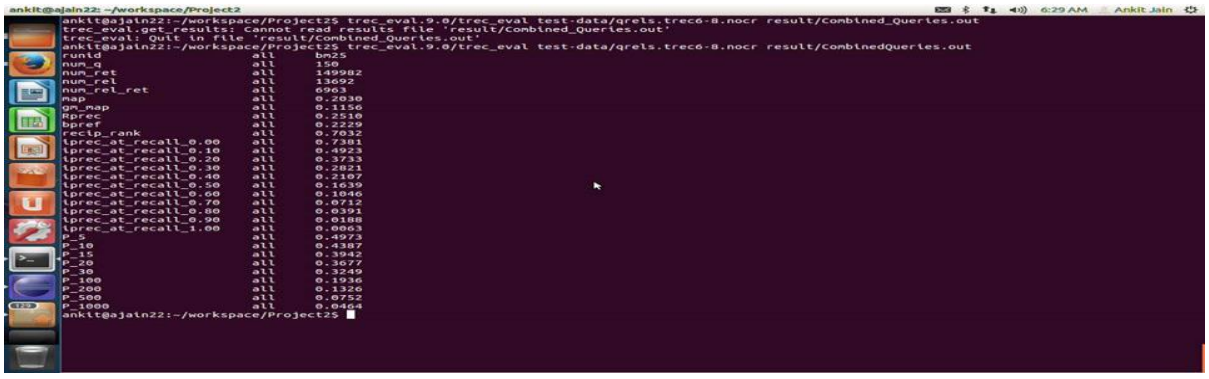
ii) Running BatchSearch.java to extract result for BM25 for above query file:



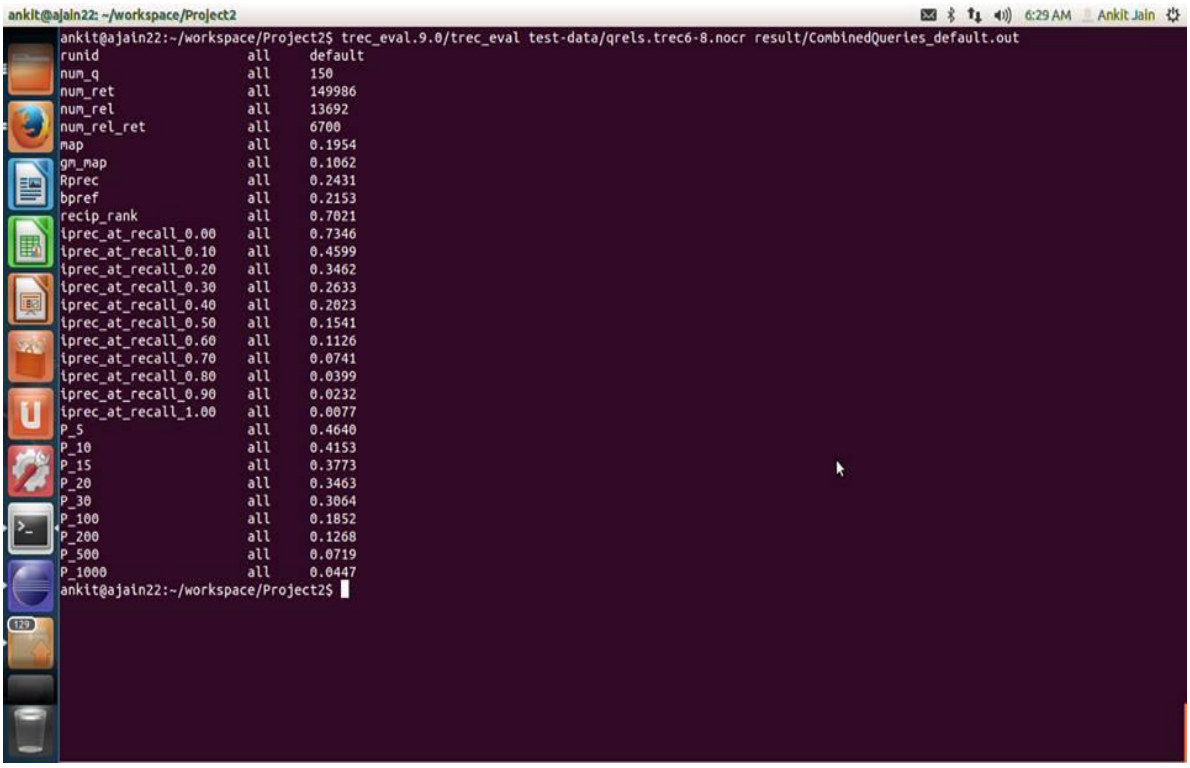
iii) Running BatchSearch.java to extract result for VSM for above query file:



iv) Running the below command for BM25 model:



v) Running the below command for VSM model:



PART 3

ANALYSIS OF QP RESULTS

COMPARISON OF RESULTS FOR DESCRIPTIVE QUERIES

SUMMARY STATISTICS

Parameters	BM25	VSM(DEFAULT)
Number of topics (num_q)	150	150
Total number of documents retrieved over all queries (num_ret)	149978	149984
Total number of relevant documents over all queries (num_rel)	13692	13692
Total number of relevant documents retrieved over all queries (num_rel_ret)	5561	5381

Recall Level Precision Averages Comparison

Recall Level Precision Averages		
Recall	**Precision of BM25**	**Precision of TF-IDF**
0	0.6376	0.5575
0.1	0.3765	0.3387
0.2	0.2817	0.2447
0.3	0.2103	0.1691
0.4	0.1628	0.1211
0.5	0.1201	0.0853
0.6	0.0777	0.0502
0.7	0.0443	0.0306
0.8	0.0231	0.0160
0.9	0.0101	0.0075
1	0.0040	0.0026

Average Precision over all Relevant Docs		

Interpolated (ircl_prn)	0.1771	0.1475

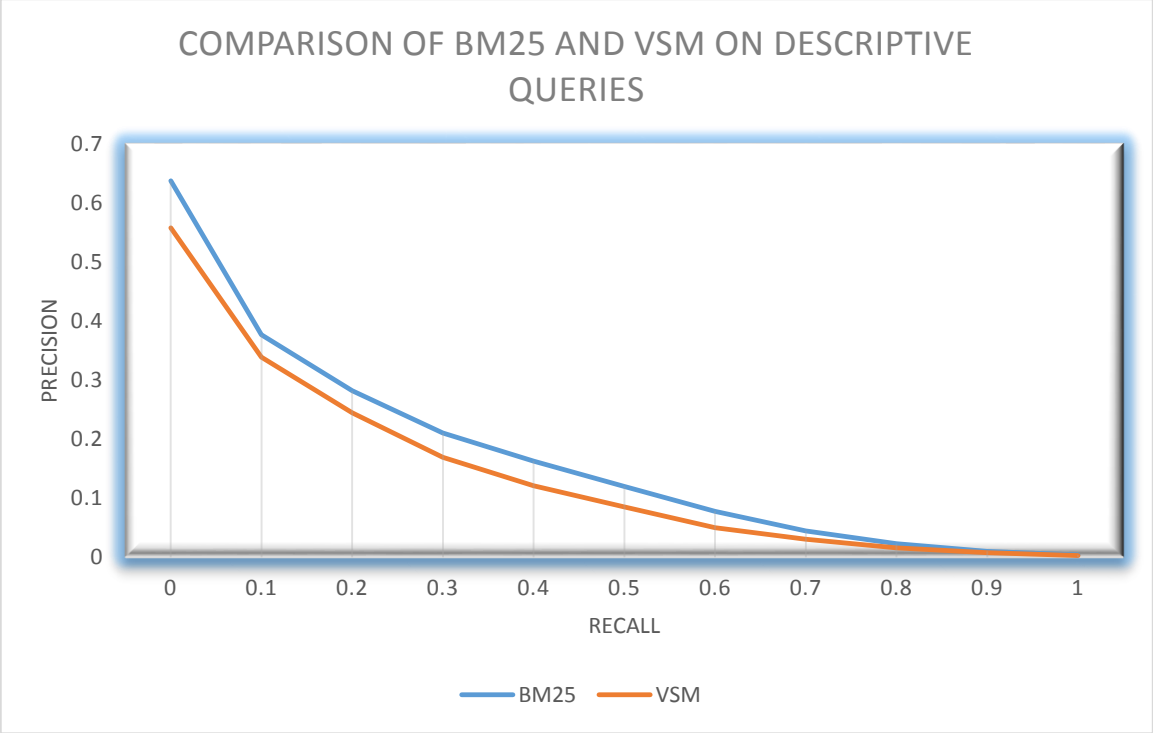
Document Level Precision Averages Comparison

Document Level Precision Averages		
No. of Docs (k)	**Precision of BM25**	**Precision of TF-IDF**
At 5 docs	0.4080	0.3573
At 10 docs	0.3540	0.3113
At 15 docs	0.3222	0.2818
At 20 docs	0.2957	0.2560
At 30 docs	0.2509	0.2222
At 100 docs	0.1527	0.1321
At 200 docs	0.1042	0.0922
At 500 docs	0.0592	0.0527
At 1000 docs	0.0371	0.0332

k-Precision (Precision after k docs Retrieved)		
Exact	0.2204	0.1932

Comparison on the basis of different measures

	BM25	**TF-IDF**
Interpolated Precision	0.1771	0.1475
Precision	0.0370	0.0358
Recall	0.4061	0.3930
F measure	0.0677	0.3746
k-precision	0.2204	0.1932
R-precision	0.2033	0.2016
MAP	0.1543	0.1514
GMAP	0.0705	0.0669
bpref	0.1814	0.1810
Reciprocal Rank	0.5961	0.6113



BM25 is better, as it is closer to the right hand corner of the curve.

COMPARISON OF RESULTS FOR NARRATIVE QUERIES

SUMMARY STATISTICS

Parameters	BM25	VSM(DEFAULT)
Number of topics (num_q)	150	150
Total number of documents retrieved over all queries (num_ret)	149976	149986
Total number of relevant documents over all queries (num_rel)	13692	13692
Total number of relevant documents retrieved over all queries (num_rel_ret)	4981	4744

Recall Level Precision Averages Comparison

Recall Level Precision Averages		
Recall	**Precision of BM25**	**Precision of TF-IDF**
0	0.5575	0.5458
0.1	0.3387	0.3108
0.2	0.2447	0.2172
0.3	0.1691	0.1418
0.4	0.1211	0.1007
0.5	0.0853	0.0692
0.6	0.0502	0.0430
0.7	0.0306	0.0243
0.8	0.0160	0.0142
0.9	0.0075	0.0058
1	0.0026	0.0031

Average Precision over all Relevant Docs		
Interpolated (ircl_prn)	0.1475	0.1341

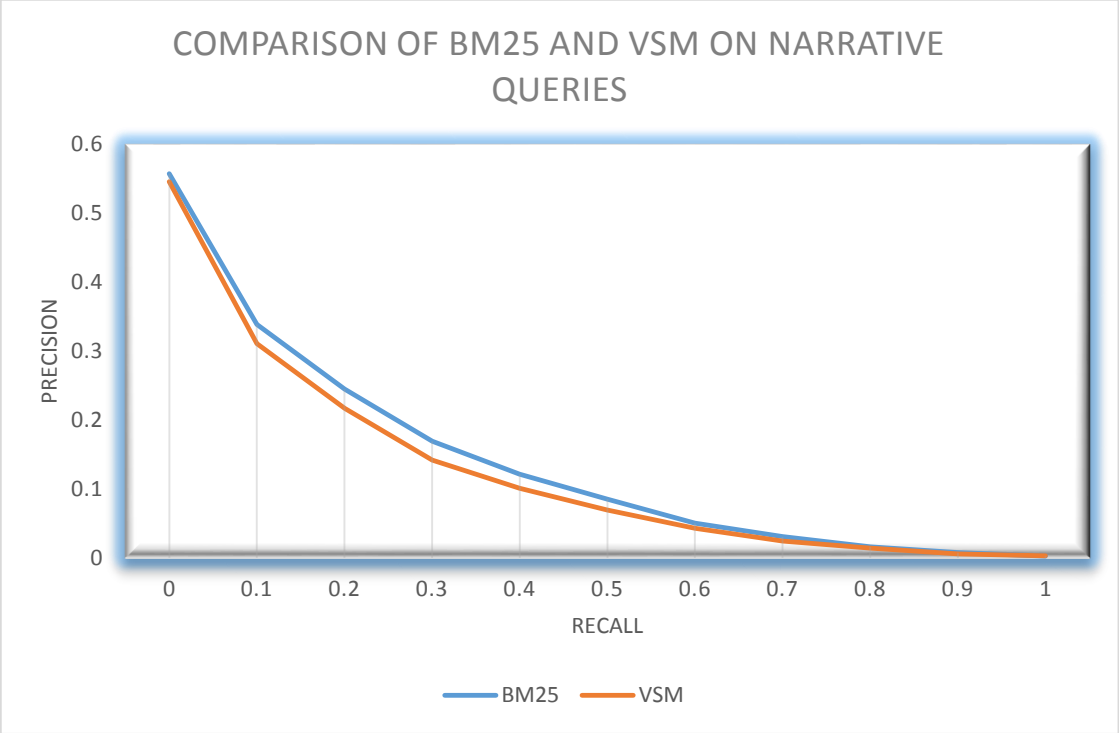
Document Level Precision Averages Comparison

Document Level Precision Averages		
No. of Docs (k)	**Precision of BM25**	**Precision of TF-IDF**
At 5 docs	0.3573	0.3347
At 10 docs	0.3113	0.2793
At 15 docs	0.2818	0.2524
At 20 docs	0.2560	0.2300
At 30 docs	0.2222	0.2029
At 100 docs	0.1321	0.1207
At 200 docs	0.0922	0.0848
At 500 docs	0.0527	0.0494
At 1000 docs	0.0332	0.0316

k-Precision (Precision after k docs Retrieved)		
Exact	0.1932	0.1762

Comparison on the basis of different measures

	BM25	**TF-IDF**
Interpolated Precision	0.1475	0.1341
Precision	0.0332	0.0316
Recall	0.3637	0.3464
F measure	0.0607	0.0576
k-precision	0.1932	0.1762
R-precision	0.1686	0.1533
MAP	0.1272	0.1127
GMAP	0.0342	0.0290
bpref	0.1654	0.1465
Reciprocal Rank	0.5290	0.5158



BM25 is better, as it is closer to the right hand corner of the curve

COMPARISON OF RESULTS FOR COMBINED QUERIES

SUMMARY STATISTICS

Parameters	BM25	VSM(DEFAULT)
Number of topics (num_q)	150	150
Total number of documents retrieved over all queries (num_ret)	149982	149986
Total number of relevant documents over all queries (num_rel)	13692	13692
Total number of relevant documents retrieved over all queries (num_rel_ret)	6963	6700

Recall Level Precision Averages Comparison

Recall Level Precision Averages		
Recall	**Precision of BM25**	**Precision of TF-IDF**
0	0.7381	0.7346
0.1	0.4923	0.4599
0.2	0.3733	0.3462
0.3	0.2821	0.2633
0.4	0.2107	0.2023
0.5	0.1639	0.1541
0.6	0.1046	0.1126
0.7	0.0712	0.0741
0.8	0.0391	0.0399
0.9	0.0188	0.0232
1	0.0063	0.0077

Average Precision over all Relevant Docs		
Interpolated (ircl_prn)	0.2273	0.2198

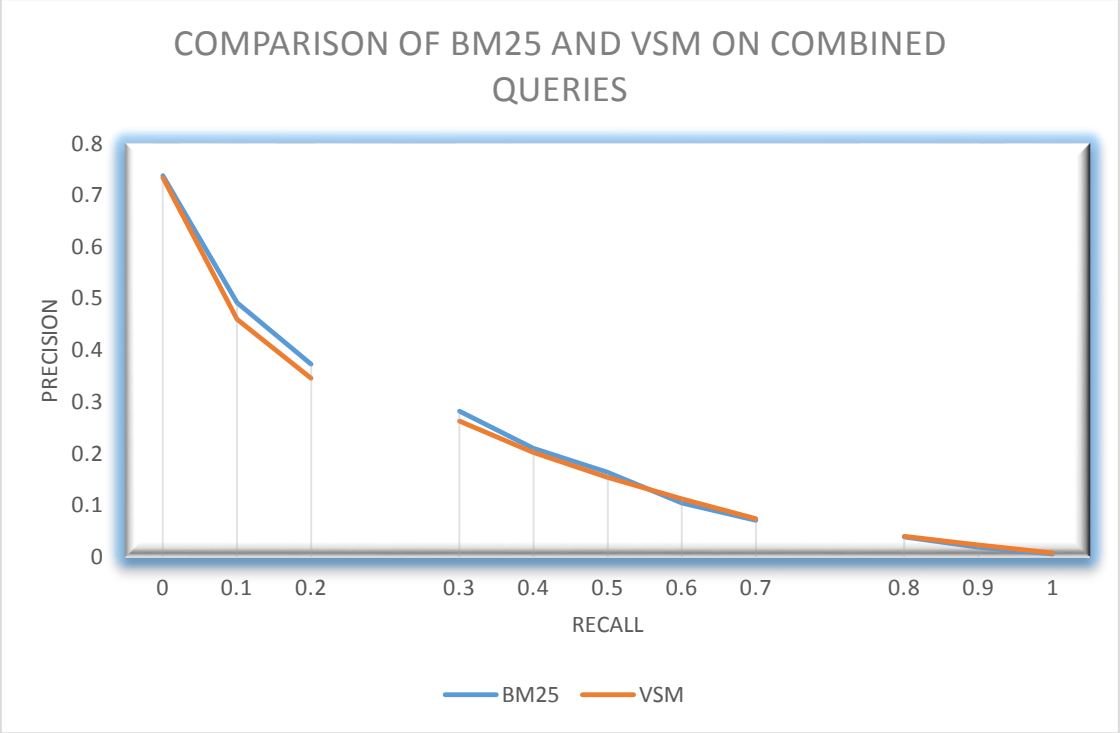
Document Level Precision Averages Comparison

Document Level Precision Averages		
No. of Docs (k)	**Precision of BM25**	**Precision of TF-IDF**
At 5 docs	0.4973	0.4640
At 10 docs	0.4387	0.4153
At 15 docs	0.3942	0.3773
At 20 docs	0.3677	0.3463
At 30 docs	0.3249	0.3064
At 100 docs	0.1936	0.1852
At 200 docs	0.1326	0.1268
At 500 docs	0.0752	0.0719
At 1000 docs	0.0464	0.0447

k-Precision (Precision after k docs Retrieved)		
Exact	0.2745	0.2597

Comparison on the basis of different measures

	BM25	**TF-IDF**
Interpolated Precision	0.2273	0.2198
Precision	0.0464	0.0446
Recall	0.5085	0.4893
F measure	0.0848	0.0816
k-precision	0.2745	0.2597
R-precision	0.2510	0.2431
MAP	0.2030	0.1954
GMAP	0.1156	0.1062
bpref	0.2229	0.2153
Reciprocal Rank	0.7032	0.7021

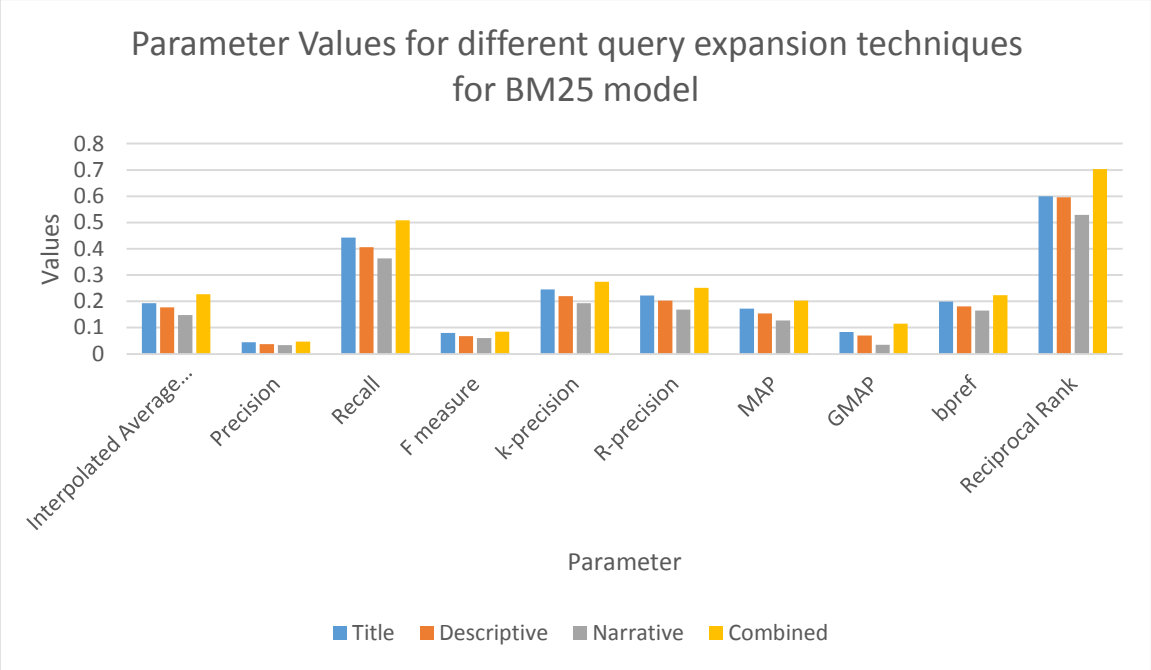


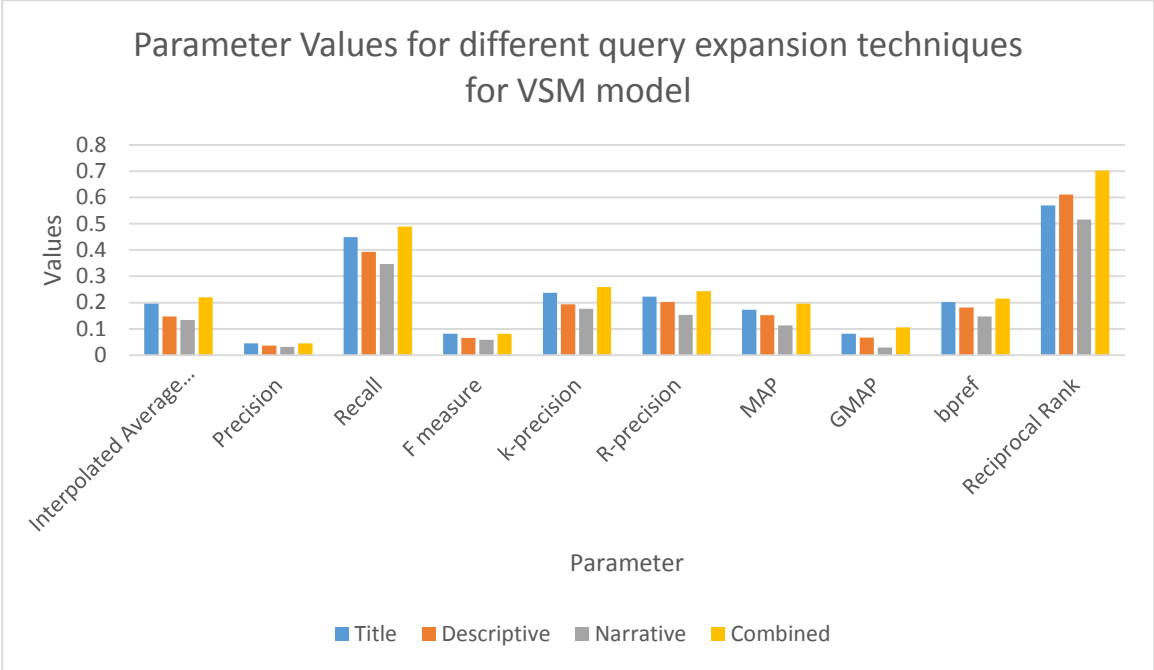
For high precision, BM25 is better.
For middle recall, BM25 is better.
For high recall performance, the two model performs equally well.

SUMMARY OF COMPARISONS

Parameter	Descriptive		Narrative		Combined		Model which performed better		
	BM25	VSM	BM25	VSM	BM25	VSM	For Desc	For Narr	For Combined
Interpolated Average Precision	0.1771	0.1475	0.1475	0.1341	0.2273	0.2198	BM25	BM25	BM25
Precision	0.037	0.0358	0.0332	0.0316	0.0464	0.0446	BM25	BM25	BM25
Recall	0.4061	0.393	0.3637	0.3464	0.5085	0.4893	BM25	BM25	BM25
F measure	0.0677	0.0656	0.0607	0.0576	0.0848	0.0816	BM25	BM25	BM25
k-precision	0.2204	0.1932	0.1932	0.1762	0.2745	0.2597	BM25	BM25	BM25
R-precision	0.2033	0.2016	0.1686	0.1533	0.251	0.2431	BM25	BM25	BM25
MAP	0.1543	0.1514	0.1272	0.1127	0.203	0.1954	BM25	BM25	BM25
GMAP	0.0705	0.0669	0.0342	0.029	0.1156	0.1062	BM25	BM25	BM25
bpref	0.1814	0.181	0.1654	0.1465	0.2229	0.2153	BM25	BM25	BM25
Reciprocal Rank	0.5961	0.6113	0.529	0.5158	0.7032	0.7021	VSM	BM25	BM25

*The columns marked in green are greater values between the two models.


Parameter Values for different query expansion techniques for BM25 model
Values
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Interpolated Average...
Precision
Recall
F measure
k-precision
R-precision
MAP
GMAP
bpref
Reciprocal Rank
Parameter
Title
Descriptive
Narrative
Combined


Parameter Values for different query expansion techniques for VSM model
Values
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Interpolated Average...
Precision
Recall
F measure
k-precision
R-precision
MAP
GMAP
bpref
Reciprocal Rank
Parameter
Title
Descriptive
Narrative
Combined

SECTION 3

1) **Okapi (BM25)** performed best overall.

2) We have used the following query expansion techniques,

i) In the first technique, query consist of words from title.

ii) In the second technique, query consist of words from description part. Stop words and non-alphanumeric characters were removed from the query.

iii) In the third technique, query consist of words from narration part. In this case also, stop words and non-alphanumeric characters were removed from the query.

iv) In the fourth technique, query consist of words from title, description and narration part. In this case also, stop words and non-alphanumeric characters were removed from the query.

The working of query expansion techniques depends on the following points:

i) In our case, the combined queries are best performing because of the highest precision values observed in this case. The high performance can be owned from the fact that in this case more number of query terms will be present in the collection, precision will improve. Net similarity score between the query and document in this case will be the highest and thus more relevant documents can be retrieved. In the combined query, chances that the query term will be present as it is in the collection are more, and so the precision achieved is also more.

3) **Other query expansion techniques that can be employed are:**

i) Query terms can be chosen from the combination of title and description part, or from the combination of title and narration part or from the combination of description and narration part.

ii) Stemming can be done for the query words. By stemming a user-entered term, more documents are matched, as the alternate word forms for a user entered term are matched as well, thereby increasing the total recall. This comes at the expense of reducing the precision.

iii) Other approaches that can be included are assigning weights to the terms from the three different zones, or picking up noun words for the query.