# End-Sem Report
# Loveleen Girdhar
# MT15029

## Solution 1:

Formula for **Rocchio Feedback** is given below:

$$\overrightarrow{Q_m} = (a \cdot \overrightarrow{Q_o}) + \left(b \cdot \frac{1}{|D_r|} \cdot \sum_{\overrightarrow{D_j} \in D_r} \overrightarrow{D_j}\right) - \left(c \cdot \frac{1}{|D_{nr}|} \cdot \sum_{\overrightarrow{D_k} \in D_{nr}} \overrightarrow{D_k}\right)$$

**Parameter** used in above given formula are alpha,beta and gamma. In above image these are shown by **a,b,c** respectively.Here **Qm** is the revised query vector and **Q0** is our initial **query vector**. **Dr** is here set of **relevant documents** taken from the user as feedback. **Dnr** here is set of **non-relevant documents** taken from user as feedback. Weights used in the above formula are responsible for reshaping the  query Qm. To move the query vector away from the centroid of relevant documents , we need to use the lower value of **beta(b)**. If we use the **lower value of beta(b)** then the relevant documents vector will not affect the original query vector much and also will not affect the query much. Vector gets dependent on non-relevant document vector The query vector can move either towards the non-relevant documents or away from both the relevant and non-relevant documents. This will be dependent on the value of gamma. If we keep a **higher value of gamma(c)**, then the query vector will move away from both relevant and non-relevant documents and if we keep a lower value of gamma. If value of **gamma(c) > beta(b)**, then the query vector will move towards the non-relevant documents.

## Solution 2:

Here we are using three parameters alpha(a), beta(b) and gamma(c). Approximate values used for alpha(a)= 1, beta(b)=0.75 and gamma(c)=1.

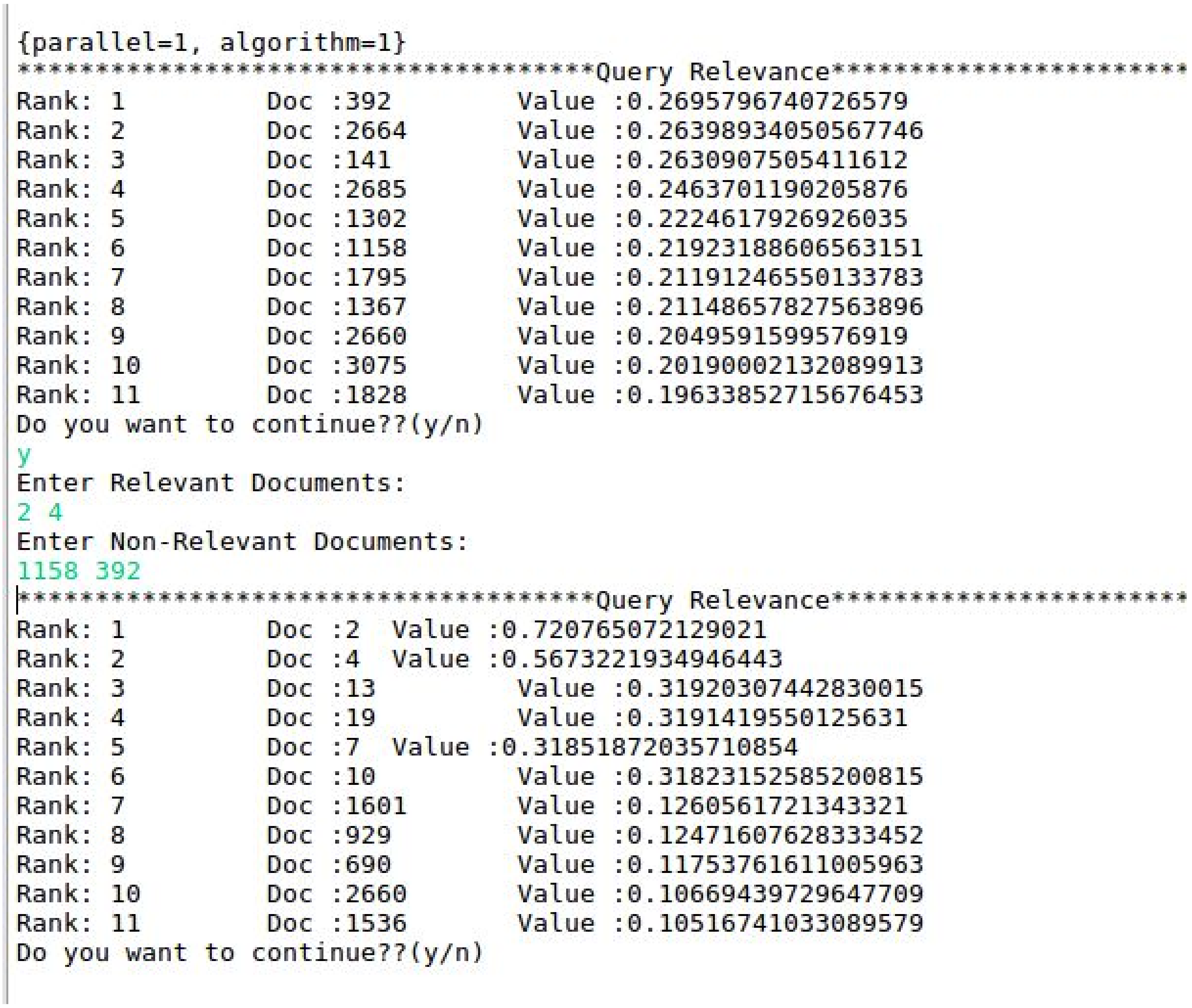
Let us suppose we have Documents returned as result by **Query Relevance Feedback.**

D={1,5,9,8} Where 1 has higher rank than 5 and so on. Now It will ask from user about relevant document and Irrelevant document.

Relevant Document ={8,9} and Irrelevant Document={1}. It will increase the score obtained by the document 8 and 9. And remove Irrelevant document 1 from the list of document returned on the basis of formation of Revised Query Vector(Qm).

**We keep value of alpha(a)= 1 to keep related document to original query. And beta (0.75 to 1 ) because of that search engine adds new results to result set. And gamma(c) =1 to nullify the weights of irrelevant terms from query vector so that it will not search again for those documents.**

Screenshot attached of running Query Relevance Feedback below result for Query Parallel Algorithms. Here we have entered **Relevant documents {2,4}** which was not even in result but added in the reformed query. And we have entered **non-relevant documents {1158,392} .** which is being removed from resultset.



```
{parallel=1, algorithm=1}
*************************************Query Relevance***********************
Rank: 1         Doc :392        Value :0.2695796740726579
Rank: 2         Doc :2664       Value :0.26398934050567746
Rank: 3         Doc :141        Value :0.2630907505411612
Rank: 4         Doc :2685       Value :0.2463701190205876
Rank: 5         Doc :1302       Value :0.2224617926926035
Rank: 6         Doc :1158       Value :0.21923188606563151
Rank: 7         Doc :1795       Value :0.21191246550133783
Rank: 8         Doc :1367       Value :0.21148657827563896
Rank: 9         Doc :2660       Value :0.2049591599576919
Rank: 10        Doc :3075       Value :0.20190002132089913
Rank: 11        Doc :1828       Value :0.19633852715676453
Do you want to continue??(y/n)
y
Enter Relevant Documents:
2 4
Enter Non-Relevant Documents:
1158 392
*************************************Query Relevance***********************
Rank: 1         Doc :2  Value :0.720765072129021
Rank: 2         Doc :4  Value :0.5673221934946443
Rank: 3         Doc :13         Value :0.31920307442830015
Rank: 4         Doc :19         Value :0.3191419550125631
Rank: 5         Doc :7  Value :0.31851872035710854
Rank: 6         Doc :10         Value :0.31823152585200815
Rank: 7         Doc :1601       Value :0.1260561721343321
Rank: 8         Doc :929        Value :0.12471607628333452
Rank: 9         Doc :690        Value :0.11753761611005963
Rank: 10        Doc :2660       Value :0.10669439729647709
Rank: 11        Doc :1536       Value :0.10516741033089579
Do you want to continue??(y/n)
```

## Solution 3:

Tf-Idf score: 1 + (tf * idf) , where tf = term frequency (tft,d) and Idf = log10(N/dft).
Score are formula dependent. Values shown below are based on above used formula. If someone is using weighted Tf-Idf formula than result may vary.
Tf-Idf Score Representation
Rank: DocumentId: Score
BM25 Reresentation
Rank: DocumentId: Score

| Query | Tf-idf score<br>(Rank#:Document:Score) | BM 25 score<br>(Rank#:Document:Score) |
|---|---|---|

| | | |
|---|---|---|
| Portable operating system | Rank1:3127:20.270170813086473<br>Rank2:1591:15.384488835622127<br>Rank3:1680:14.73552521859281<br>Rank4:2319:13.437597984534175<br>Rank5:2740:12.788634367504859<br>Rank6:1930:12.226890012304725<br>Rank7:2379:11.571956174110735<br>Rank8:2246:11.352988414440265<br>Rank9:1844:10.734454255439761<br>Rank10:3068:10.55788622823450 | Rank1:3127:6.20285665664153<br>Rank2:2246:4.605462038205922<br>Rank3:1930:3.8571192345558725<br>Rank4:3196:3.64136906761334<br>Rank5:2593:2.663626850499866<br>Rank6:3068:2.302397463860371<br>Rank7:2319:2.2847499995400975<br>Rank8:2740:2.2832824759938575<br>Rank9:1680:2.236597451428784<br>Rank10:2379:2.2164075268664583 |
| Parallel algorithm | Rank1:2714:15.199497123043741<br>Rank2:1811:12.295185333052038<br>Rank3:2433:10.82971423960528<br>Rank4:2289:10.358703408878583<br>Rank5:2973:9.596978692881901<br>Rank6:950:9.21611633488356<br>Rank7:2342:9.067989523608599<br>Rank8:2851:9.067989523608599<br>Rank9:2785:8.83525397688522<br>Rank10:1601:8.83525397688522 | Rank1:2714:2.800166553661334<br>Rank2:2973:2.6813146680706152<br>Rank3:2433:2.6016607722484544<br>Rank4:2664:2.589886627711763<br>Rank5:2785:2.58950569730061<br>Rank6:950:2.5615925390719227<br>Rank7:2266:2.5014498060234835<br>Rank8:1262:2.4632507506145846<br>Rank9:3075:2.459850357301431<br>Rank10:2685:2.448579434831288 |
| Applied stochastic process | Rank1:2065:11.963477610878012<br>Rank2:1696:10.706045045458302<br>Rank3:2342:9.544144934140562<br>Rank4:3043:7.8353159473124485<br>Rank5:2080:7.8353159473124485<br>Rank6:2999:7.8353159473124485<br>Rank7:3120:7.691405143807732<br>Rank8:1410:6.989754741606744<br>Rank9:2535:6.989754741606744<br>Rank10:1359:6.836990650393676 | Rank1:1696:3.7368790730707593<br>Rank2:1410:2.9899302949096844<br>Rank3:268:2.97892703732016<br>Rank4:2535:2.6995690177219074<br>Rank5:2882:2.6192599378954955<br>Rank6:1194:2.5735382715862913<br>Rank7:3020:2.511158462668063<br>Rank8:1233:2.413700021896501<br>Rank9:2065:2.37582669014344<br>Rank10:1892:2.3754486124453433 |
| Perform evaluation and model of computer system | Rank1:3048:32.8305579632387<br>Rank2:2318:31.232113269817766<br>Rank3:3070:28.276431829346688<br>Rank4:2344:27.897698801693128<br>Rank5:2542:27.39002034157057<br>Rank6:1827:26.545621599737217<br>Rank7:2319:26.53160752975533<br>Rank8:1844:26.22166409852393 | Rank1:2318:7.019112416340891<br>Rank2:3048:5.978885863012449<br>Rank3:3089:5.085566425366673<br>Rank4:3070:5.085430739620256<br>Rank5:2319:4.966707412484976<br>Rank6:2542:4.911945607096959<br>Rank7:3119:4.888623462942207<br>Rank8:2741:4.888351483430168 |

| | | |
|---|---|---|
| | Rank9:1680:25.583979578109997<br>Rank10:2188:25.0255182906044 | Rank9:2452:4.7639402730983935<br>Rank10:2894:4.610588012494716 |
| Parallel process in information retrieval | Rank1:2342:22.6870684905253<br>Rank2:1699:20.780119733698733<br>Rank3:3134:19.76617846271714<br>Rank4:2288:18.719126971904185<br>Rank5:1681:17.36293518377767<br>Rank6:2307:17.350232385203466<br>Rank7:1457:16.845093957494612<br>Rank8:1846:16.56323526790699<br>Rank9:2882:16.35119089030421<br>Rank10:2714:16.19884829478718 | Rank1:2114:5.031641766112655<br>Rank2:2967:4.691785045872723<br>Rank3:2307:4.410641635124168<br>Rank4:1457:4.38155308212249<br>Rank5:1927:4.308077879987508<br>Rank6:1601:4.2963962372426<br>Rank7:1846:4.296144466344407<br>Rank8:2342:4.2228260542533205<br>Rank9:2519:3.916835580508854<br>Rank10:1959:3.8883129390675713 |