

We have multiple relevance measures to determine whether a document is relevant or not.

We can use **LMJM** and DFR to calculate score of the documents in our corpus

Query = Halifax explosion

Below are results in trec eval format from DFR and LMJM

```
vigy@Vigneshs-MacBook-Pro code % python3 QueryProgram.py my_corpus_1.txt
```

```
"halifax explosion" DFR
```

```
1 0 7 1 0.24944012687112654 DFR
1 0 9 2 0.24944012687112654 DFR
1 0 11 3 0.2462808755877967 DFR
1 0 5 4 0.2456327477391898 DFR
1 0 10 5 0.24421880899346207 DFR
1 0 3 6 0.24256864199393174 DFR
1 0 1 7 0.24244262955669588 DFR
1 0 2 8 0.2409406323662104 DFR
1 0 4 9 0.24081630582293934 DFR
1 0 6 10 0.2373865286549363 DFR
1 0 8 11 0.08017109122459541 DFR
```

```
vigy@Vigneshs-MacBook-Pro code % python3 QueryProgram.py my_corpus_1.txt
```

```
"halifax explosion" LMJM
```

```
1 0 6 1 74.84100201578056 LMJM
1 0 4 2 72.58981524929995 LMJM
1 0 3 3 72.43998884819356 LMJM
1 0 2 4 70.72260155971225 LMJM
1 0 1 5 68.89549469364628 LMJM
1 0 5 6 68.5943442873523 LMJM
1 0 10 7 66.92482993580322 LMJM
1 0 7 8 66.73479026956335 LMJM
1 0 11 9 66.13212226896721 LMJM
1 0 9 10 64.53756582622665 LMJM
1 0 8 11 32.813971337101044 LMJM
```

Using trec eval tool and comparing results from both:

Trec eval for query 1 for DFR

```
[vigy@Vigneshs-MacBook-Pro files % trec_eval trec_rel_query_1.txt trec_rel_top_DFR_query_1.txt
runid          all      DFR
num_q          all      1
num_ret        all      11
num_rel        all      10
num_rel_ret    all      10
map            all      1.0000
gm_map         all      1.0000
Rprec          all      1.0000
bpref          all      1.0000
recip_rank     all      1.0000
iprec_at_recall_0.00 all 1.0000
iprec_at_recall_0.10 all 1.0000
iprec_at_recall_0.20 all 1.0000
iprec_at_recall_0.30 all 1.0000
iprec_at_recall_0.40 all 1.0000
iprec_at_recall_0.50 all 1.0000
iprec_at_recall_0.60 all 1.0000
iprec_at_recall_0.70 all 1.0000
iprec_at_recall_0.80 all 1.0000
iprec_at_recall_0.90 all 1.0000
iprec_at_recall_1.00 all 1.0000
P_5            all      1.0000
P_10           all      1.0000
P_15           all      0.6667
P_20           all      0.5000
P_30           all      0.3333
P_100          all      0.1000
P_200          all      0.0500
P_500          all      0.0200
P_1000         all      0.0100
```

Trec eval for query 1 for LMJM

```
[vigy@Vigneshs-MacBook-Pro files % trec_eval trec_rel_query_1.txt trec_rel_top_LMJM_query_1.txt
runid          all      LMJM
num_q          all      1
num_ret        all      11
num_rel        all      10
num_rel_ret    all      10
map            all      1.0000
gm_map         all      1.0000
Rprec          all      1.0000
bpref          all      1.0000
recip_rank     all      1.0000
iprec_at_recall_0.00 all 1.0000
iprec_at_recall_0.10 all 1.0000
iprec_at_recall_0.20 all 1.0000
iprec_at_recall_0.30 all 1.0000
iprec_at_recall_0.40 all 1.0000
iprec_at_recall_0.50 all 1.0000
iprec_at_recall_0.60 all 1.0000
iprec_at_recall_0.70 all 1.0000
iprec_at_recall_0.80 all 1.0000
iprec_at_recall_0.90 all 1.0000
iprec_at_recall_1.00 all 1.0000
P_5            all      1.0000
P_10           all      1.0000
P_15           all      0.6667
P_20           all      0.5000
P_30           all      0.3333
P_100          all      0.1000
P_200          all      0.0500
P_500          all      0.0200
P_1000         all      0.0100
```

Query = huge casualties

vigy@Vigneshs-MacBook-Pro code % python3 QueryProgram.py my_corpus_1.txt
"huge casualties" DFR

1 0 8 1 1.5696151079271115 DFR
1 0 5 2 1.5240334537782334 DFR
1 0 4 3 1.4197661447188794 DFR
1 0 3 4 0.6127531616325455 DFR
1 0 2 5 0.6004998121550662 DFR

vigy@Vigneshs-MacBook-Pro code % python3 QueryProgram.py my_corpus_1.txt
"huge casualties" LMJM

1 0 4 1 61.71365738639936 LMJM
1 0 5 2 60.12613232513636 LMJM
1 0 8 3 59.27183621205735 LMJM
1 0 2 4 31.244977597284617 LMJM
1 0 3 5 31.005059008968665 LMJM

Using trec eval tool and comparing results from both:

Trec eval for query 2 for DFR

```
[vigy@Vigneshs-MacBook-Pro files % trec_eval trec_rel_query_2.txt trec_rel_top_DFR_query_2.txt
runid          all      DFR
num_q          all      1
num_ret        all      5
num_rel        all      6
num_rel_ret    all      5
map            all      0.8333
gm_map         all      0.8333
Rprec          all      0.8333
bpref          all      0.8333
recip_rank     all      1.0000
iprec_at_recall_0.00 all    1.0000
iprec_at_recall_0.10 all    1.0000
iprec_at_recall_0.20 all    1.0000
iprec_at_recall_0.30 all    1.0000
iprec_at_recall_0.40 all    1.0000
iprec_at_recall_0.50 all    1.0000
iprec_at_recall_0.60 all    1.0000
iprec_at_recall_0.70 all    1.0000
iprec_at_recall_0.80 all    1.0000
iprec_at_recall_0.90 all    0.0000
iprec_at_recall_1.00 all    0.0000
P_5            all      1.0000
P_10           all      0.5000
P_15           all      0.3333
P_20           all      0.2500
P_30           all      0.1667
P_100          all      0.0500
P_200          all      0.0250
P_500          all      0.0100
P_1000         all      0.0050
```

Trec eval for query 2 for LMJM

```

[vigy@Vigneshs-MacBook-Pro files % trec_eval trec_rel_query_2.txt trec_rel_top_LMJM_query_2.txt
runid          all      LMJM
num_q          all      1
num_ret        all      5
num_rel        all      6
num_rel_ret    all      5
map            all      0.8333
gm_map         all      0.8333
Rprec          all      0.8333
bpref          all      0.8333
recip_rank     all      1.0000
iprec_at_recall_0.00 all    1.0000
iprec_at_recall_0.10 all    1.0000
iprec_at_recall_0.20 all    1.0000
iprec_at_recall_0.30 all    1.0000
iprec_at_recall_0.40 all    1.0000
iprec_at_recall_0.50 all    1.0000
iprec_at_recall_0.60 all    1.0000
iprec_at_recall_0.70 all    1.0000
iprec_at_recall_0.80 all    1.0000
iprec_at_recall_0.90 all    0.0000
iprec_at_recall_1.00 all    0.0000
P_5            all      1.0000
P_10           all      0.5000
P_15           all      0.3333
P_20           all      0.2500
P_30           all      0.1667
P_100          all      0.0500
P_200          all      0.0250
P_500          all      0.0100
P_1000         all      0.0050

```

Because the dataset was small, the DFR and LMJM relevance calculation measure are relatively same.