Sneha Panicker
UBIT name- spanicke
Person Number- 50316848

Project 3: Evaluation of IR Models

Introduction:

In this project we have implemented and evaluated 3 Information retrieval models for the IR system and tried to improve the search results based on the understanding of the models. Twitter data in languages English, German and Russian along with their queries and corresponding judgement values are given and this data is indexed into Solr to evaluate 3 models: 1) BM25, 2) Language Model, 3) Divergence from Randomness (DFR). The results from these models will be evaluated using Trec_eval program. Based on these results the performance is improved in terms of Mean Average Precision (MAP).

MAP Values of IR Models with Default Configurations:

1. BM 25 Model: 0.6675
2. Language Model: 0.6639
3. Divergence from Randomness Model: 0.6722

Improvements on the IR Models:

BM25

- Synonyms for words that occur in queries were added to the synonyms.txt in the conf folder. This help to increase the relevance score of the tweets which have synonyms to the words present in the queries.
- Instead of using standard query parser we have used Dismax quey parser to check results.
- b = 0.5 and k = 1.1, since length of the tweet does not affect the relevance drastically since tweets are length restricted to a small range.
- 0.7019 is the maximum MAP obtained.

Schema

```
<similarity class="solr.BM25SimilarityFactory">
  <float name="b">0.5</float>
  <float name="k1">1.1</float>
</similarity>
```

```
runid                   all     bm25
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     141
map                     all     0.7019
gm_map                  all     0.6686
Rprec                   all     0.6832
bpref                   all     0.7218
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     0.9815
iprec_at_recall_0.30    all     0.8726
iprec_at_recall_0.40    all     0.8426
iprec_at_recall_0.50    all     0.8111
iprec_at_recall_0.60    all     0.6534
iprec_at_recall_0.70    all     0.5719
iprec_at_recall_0.80    all     0.4247
iprec_at_recall_0.90    all     0.3226
iprec_at_recall_1.00    all     0.3226
P_5                     all     0.8267
P_10                    all     0.6800
P_15                    all     0.5467
P_20                    all     0.4700
P_30                    all     0.3133
P_100                   all     0.0940
P_200                   all     0.0470
P_500                   all     0.0188
P_1000                  all     0.0094
```

Language Model

- Synonyms for words that occur in queries were added to the synonyms.txt in the conf folder. This help to increase the relevance score of the tweets which have synonyms to the words present in the queries.
- Instead of using standard query parser we have used Dismax quey parser to check results.
- 0.6325 is the maximum MAP obtained.

Schema

```
<similarity class="solr.LMDirichletSimilarityFactory">
</similarity>
```

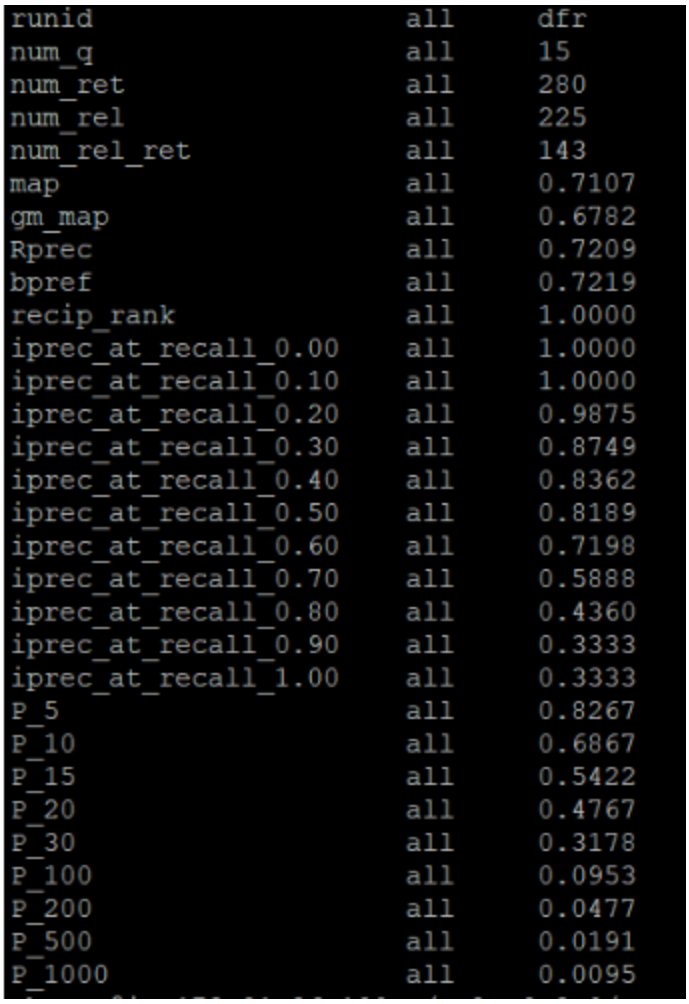
```
runid                   all     lm
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     125
map                     all     0.6325
gm_map                  all     0.5655
Rprec                   all     0.6501
bpref                   all     0.6722
recip_rank              all     1.0000
iprec_at_recall_0.00    all     1.0000
iprec_at_recall_0.10    all     0.9800
iprec_at_recall_0.20    all     0.9533
iprec_at_recall_0.30    all     0.8020
iprec_at_recall_0.40    all     0.7643
iprec_at_recall_0.50    all     0.6962
iprec_at_recall_0.60    all     0.5239
iprec_at_recall_0.70    all     0.5205
iprec_at_recall_0.80    all     0.4078
iprec_at_recall_0.90    all     0.2857
iprec_at_recall_1.00    all     0.2857
P_5                     all     0.7467
P_10                    all     0.5667
P_15                    all     0.4844
P_20                    all     0.4167
P_30                    all     0.2778
P_100                   all     0.0833
P_200                   all     0.0417
P_500                   all     0.0167
P_1000                  all     0.0083
```

Divergence from Randomness (DFR).

- Synonyms for words that occur in queries were added to the synonyms.txt in the conf folder. This help to increase the relevance score of the tweets which have synonyms to the words present in the queries.
- Instead of using standard query parser we have used Dismax quey parser to check results.
- 0.7107 is the maximum MAP obtained.

Schema

```
<similarity class="solr.DFRSimilarityFactory">
  <str name="basicModel">G</str>
  <str name="afterEffect">B</str>
  <str name="normalization">H2</str>
  <float name="c">7</float>
</similarity>
```



runid	all	dfr
num_q	all	15
num_ret	all	280
num_rel	all	225
num_rel_ret	all	143
map	all	0.7107
gm_map	all	0.6782
Rprec	all	0.7209
bpref	all	0.7219
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	0.9875
iprec_at_recall_0.30	all	0.8749
iprec_at_recall_0.40	all	0.8362
iprec_at_recall_0.50	all	0.8189
iprec_at_recall_0.60	all	0.7198
iprec_at_recall_0.70	all	0.5888
iprec_at_recall_0.80	all	0.4360
iprec_at_recall_0.90	all	0.3333
iprec_at_recall_1.00	all	0.3333
P_5	all	0.8267
P_10	all	0.6867
P_15	all	0.5422
P_20	all	0.4767
P_30	all	0.3178
P_100	all	0.0953
P_200	all	0.0477
P_500	all	0.0191
P_1000	all	0.0095

Improving BM25 Model:

After altering hyperparameters b and K1 for BM25 model, MAP score was increased.

	b	k1	MAP Score
Default values	1.2	0.75	0.6675
After Improving	1.4	1.8	0.6998
	0.75	1.1	0.7019

Conclusion:

Performance of IR models is improved by using Synonyms for query expansion which improved the relevance score of tweets. From observation, Divergence from Randomness Model (DFR) has the highest map value of 0.7107