# Information Retrieval – Project 3 Report
# Evaluation of IR Models

**Gulfam Hussain**
**UBIT Name: gulfamhu**
Department of Computer Science
University at Buffalo, NY 14214
gulfamhu@buffalo.edu

**Overview:**

As part of this projects, the aim is to implement various IR models, evaluate the IR system and improve the search results based on the understanding of the models, the implementation and the evaluation.

Twitter data in three languages - English, German and Russian, with 15 sample queries and the corresponding relevance judgements are given. We will index the given twitter data using Solr and implement the following three IR models:

1. Language Model,
2. BM25 and
3. Divergence from Randomness (DFR) Model.

The results from these three sets will be evaluated using the Trec_eval program. Based on the evaluation results, we are required to attempt to improve the performance in terms of Mean Average Precision (MAP).

**Dataset:**

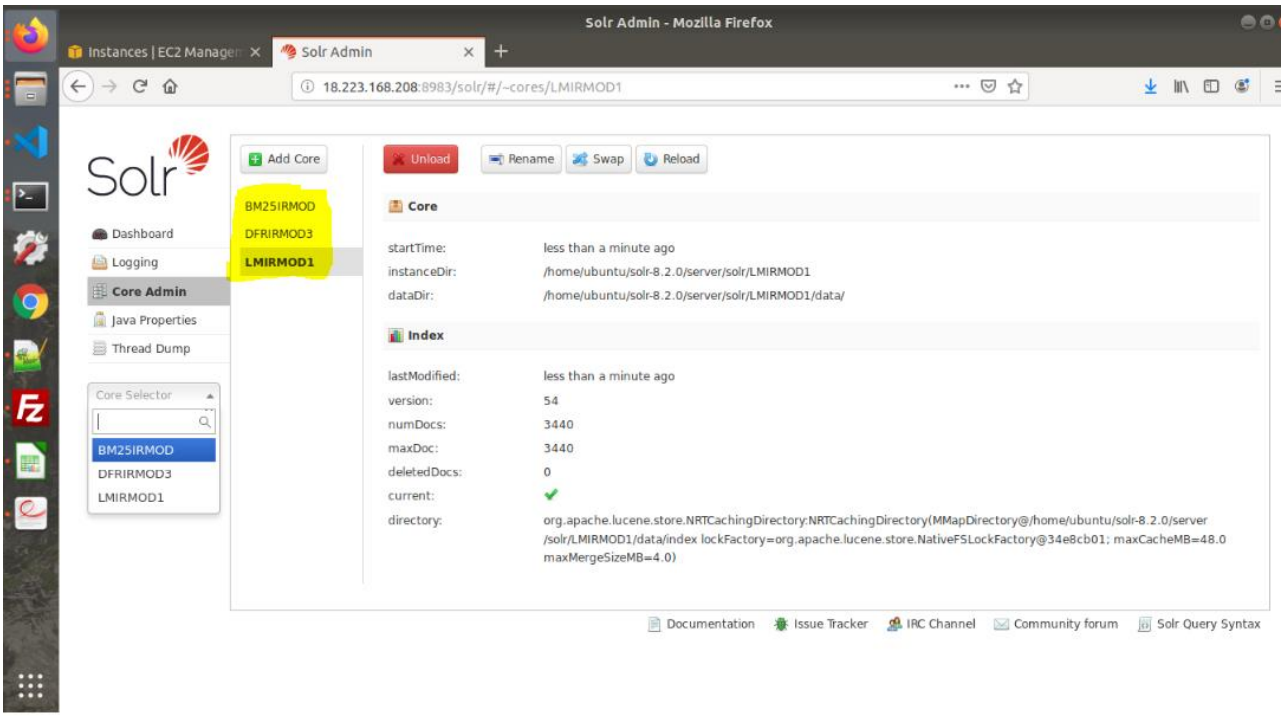
1. training_tweet.json file has been provided.
2. The other required data and details have also been provided.

**Implementation of IR Models:**

The 3 different IR models have been implemented in the Solr. Below steps are followed to create 3 different cores corresponding to each IR models with default schema in the solr.

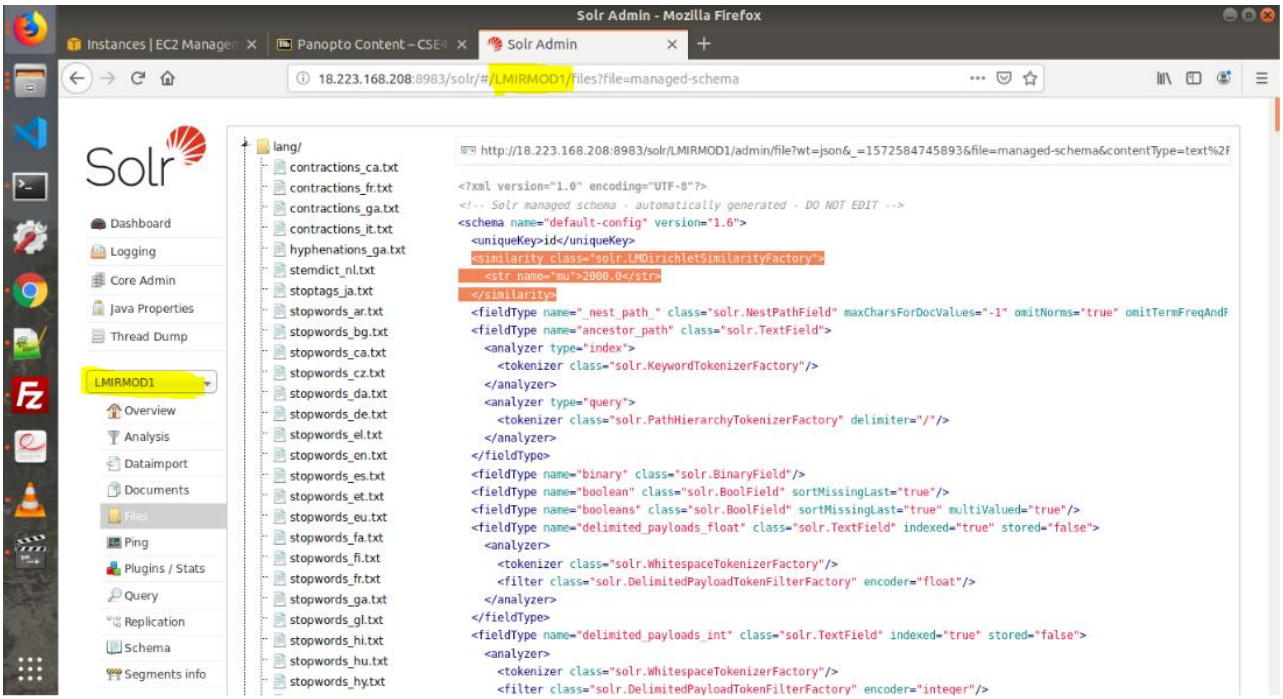
a) Created 3 different cores in the solr for each of the individual IR models.
b) The default schema along with the default values of hyperparameters were used while creating the models at the first time.
c) The given training_tweet.json dataset was posted to each of the different models in the solr.
d) Based on the given set of training queries, the output files were generated in the Trec_eval format as mentioned in the project requirement documents.
e) The trec_eval was run and the MAP(all) scores were obtained for each of the models separately on the training queries.
f) Based on the MAP (all) scores for each of the IR models on default values, I tweaked few of the hyper-parameters values to achieve the best MAP(all) scores for each of the models. The observations are listed below in this report.

# Information Retrieval – Project 3 Report
# Evaluation of IR Models



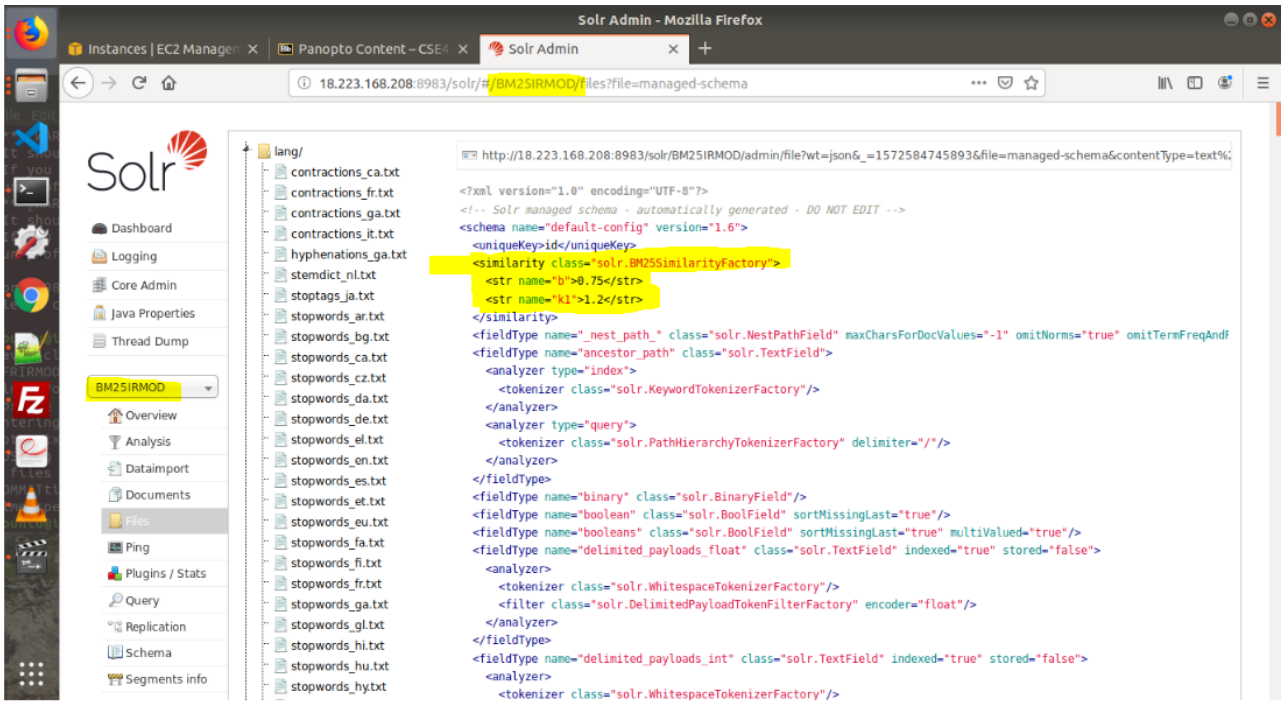

## IR models with default parameters:

### 1) Language Model (LM):

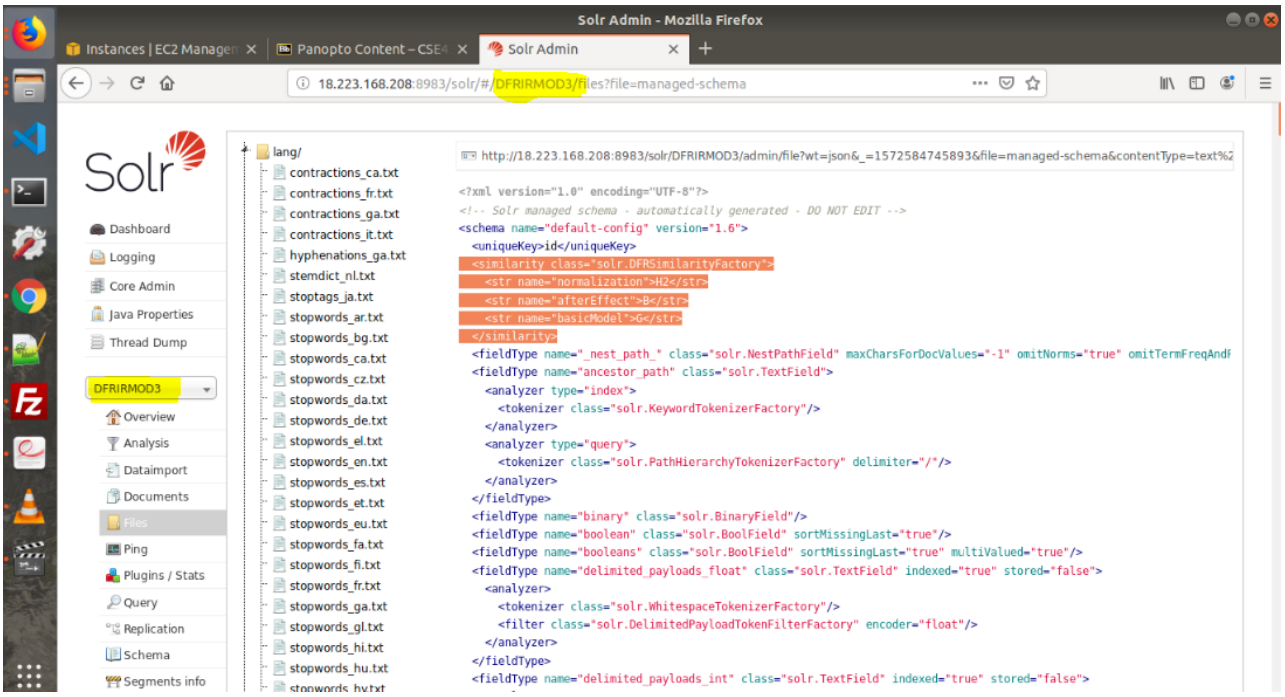


# Information Retrieval – Project 3 Report
# Evaluation of IR Models

## 2) BM25:




## 3) Divergence from Randomness (DFR) Model:



# Information Retrieval – Project 3 Report
# Evaluation of IR Models

**Improving the IR systems:**

## 1) BM25:

There are two hyper-parameters available for BM25 model which are mainly “k1” and “b” values. As per understanding, the parameter “b” controls the length of the document and the parameter “k1” controls the frequency of the word.

Frequency is actually related to the length of the document and thus, if the document is large/small, the k1 value is increased/decreased to achieve good MAP score.

As per the same logic, I tried to tweak these two parameters values as shown below and obtained the MAP(all) scores respectively and compared the new MAP score with the MAP score obtained from the default values of these hyper-parameters. As shown below in the table, the MAP score corresponding to k1=1.1 & b=0.8 was good as compared to rest of the values.

| Model | Default Hyper-parameters | Default MAP(all) score | Tweaked Hyper-parameters values | New MAP(all) score |
|---|---|---|---|---|
| BM25 | <str name="b">0.75</str><br><str name="k1">1.2</str> | 0.6756 | <str name="b">0.8</str><br><str name="k1">1.5</str> | 0.6748 |
| | | | <str name="b">0.5</str><br><str name="k1">1.7</str> | 0.6758 |
| | | | <str name="b">0.75</str><br><str name="k1">1.1</str> | 0.6758 |
| | | | <str name="b">0.5</str><br><str name="k1">1</str> | 0.6747 |
| | | | <str name="b">0.7</str><br><str name="k1">2</str> | 0.6733 |
| | | | <str name="b">0.8</str><br><str name="k1">1.1</str> | 0.6764 |

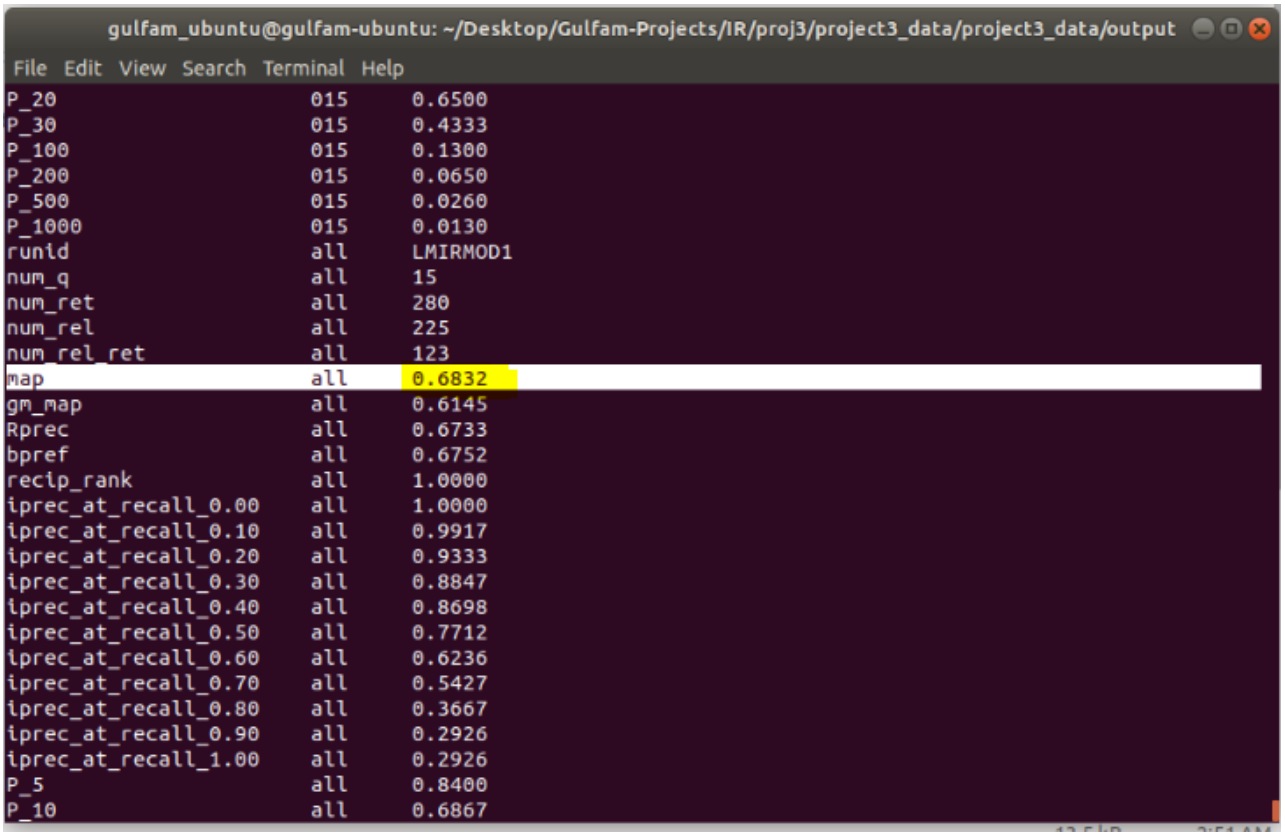
```
gulfam_ubuntu@gulfam-ubuntu: ~/Desktop/Gulfam-Projects/IR/proj3/project3_data/project3_data/output
File  Edit  View  Search  Terminal  Help
P_20                    015     0.6500
P_30                    015     0.4333
P_100                   015     0.1300
P_200                   015     0.0650
P_500                   015     0.0260
P_1000                  015     0.0130
runid                   all     BM25IRMOD
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     121
map                     all     0.6764
gm_map                  all     0.6089
Rprec                   all     0.6474
bpref                   all     0.6712
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     0.9333
iprec_at_recall_0.30    all     0.8847
iprec_at_recall_0.40    all     0.8698
iprec_at_recall_0.50    all     0.7365
iprec_at_recall_0.60    all     0.6147
iprec_at_recall_0.70    all     0.5322
iprec_at_recall_0.80    all     0.3667
iprec_at_recall_0.90    all     0.2815
iprec_at_recall_1.00    all     0.2815
P_5                     all     0.8400
P_10                    all     0.6800
```

# Information Retrieval – Project 3 Report
# Evaluation of IR Models

## 2) LM:

There is a hyper-parameter called “mu” available for LM model. The default value of “mu” is used as 2000 and the corresponding default MAP(all) score 0.6135 was achieved. After analyzing and tweaking this parameter as shown in the below table, the good MAP(all) score 0.6832 was achieved at the “mu” value of 10. Increasing the “mu” value decreased the overall MAP score. Thus it depends on the document length and the frequency of the words. For a very dense document, may be the default value work fine but for smaller document, the small mu value can be considered to achieve good overall MAP score.

| Model | Default Hyper-parameters | Default MAP(all) score | Tweaked Hyper-parameters values | New MAP(all) score |
|---|---|---|---|---|
| LM | <float name="mu">2000.0</float> | 0.6135 | <float name="mu">2200.0</float> | 0.6105 |
| | | | <float name="mu">1800.0</float> | 0.6142 |
| | | | <float name="mu">1500.0</float> | 0.6178 |
| | | | <float name="mu">1000.0</float> | 0.6297 |
| | | | <float name="mu">300.0</float> | 0.6716 |
| | | | <float name="mu">50.0</float> | 0.6813 |
| | | | <float name="mu">10.0</float> | 0.6832 |



# Information Retrieval – Project 3 Report
# Evaluation of IR Models

## 3) DFR:

There is a hyper-parameter called “c” available for DFR model along with other parameters. The default value of “c” is used as 7 in most of the cases. After analyzing and tweaking this parameter as shown in the below table, the good MAP(all) score 0.6779 was achieved at the “c” value of 2.5. Decreasing the “c” value up to 1 increases the MAP score initially and got saturated at this point.

| Model | Default Hyper-parameters | Default MAP(all) score | Tweaked Hyper-parameters values | New MAP(all) score |
|---|---|---|---|---|
| DFR | <str name="normalization">H2</str><br><str name="afterEffect">B</str><br><str name="basicModel">G</str> | 0.675 | <str name="normalization">H2</str><br><str name="afterEffect">B</str><br><str name="basicModel">G</str><br><float name="c">5</float> | 0.6771 |
| | | | float name="c">7.0</float><br><str name="normalization">H2</str><br><str name="afterEffect">B</str><br><str name="basicModel">G</str>> | 0.677 |
| | | | <float name="c">2.0</float><br><str name="normalization">H2</str><br><str name="afterEffect">B</str><br><str name="basicModel">G</str> | 0.6776 |
| | | | <float name="c">1.0</float><br><str name="normalization">H2</str><br><str name="afterEffect">B</str><br><str name="basicModel">G</str> | 0.675 |
| | | | <float name="c">3.0</float><br><str name="normalization">H2</str><br><str name="afterEffect">B</str><br><str name="basicModel">G</str> | 0.6775 |
| | | | <float name="c">2.5</float><br><str name="normalization">H2</str><br><str name="afterEffect">B</str><br><str name="basicModel">G</str> | 0.6779 |

```
gulfam_ubuntu@gulfam-ubuntu: ~/Desktop/Gulfam-Projects/IR/proj3/project3_data/project3_data/output
File  Edit  View  Search  Terminal  Help
P_20                    015     0.6500
P_30                    015     0.4333
P_100                   015     0.1300
P_200                   015     0.0650
P_500                   015     0.0260
P_1000                  015     0.0130
runid                   all     DFRIRMOD3
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     121
map                     all     0.6779
gm_map                  all     0.6066
Rprec                   all     0.6710
bpref                   all     0.6739
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     0.9333
iprec_at_recall_0.30    all     0.8772
iprec_at_recall_0.40    all     0.8603
iprec_at_recall_0.50    all     0.7363
iprec_at_recall_0.60    all     0.6211
iprec_at_recall_0.70    all     0.5459
iprec_at_recall_0.80    all     0.3667
iprec_at_recall_0.90    all     0.2926
iprec_at_recall_1.00    all     0.2926
P_5                     all     0.8267
P_10                    all     0.6867
```

# Information Retrieval – Project 3 Report
# Evaluation of IR Models

**Conclusion:**

All the 3 different IR models were implemented successfully. Also, the improvements were done on each of the models separately by tweaking the different hyper-parameters and other factors and an increase in the overall MAP scores were observed which had been listed in this report.

**References:**


1. Class slides and project description documents
2. http://lucene.apache.org/solr/7_0_0/solr-core/org/apache/solr/search/similarities/package-summary.html
3. http://lucene.apache.org/solr/guide/7_5/other-schema-elements.html#OtherSchemaElements-Similarity
4. https://lucene.apache.org/solr/8_1_0//solrcore/org/apache/solr/search/similarities/LMDirichletSimilarityFactory.html
5. http://wiki.apache.org/solr/SchemaXml#Similarity
6. http://trec.nist.gov/trec_eval/