

---

# Evaluation of IR Models

---

**Bhakti Jadhav**

Department of Computer Science  
University at Buffalo  
Buffalo NY, 14214  
[bhaktiha@buffalo.edu](mailto:bhaktiha@buffalo.edu)

## Abstract

The purpose of this project is to implement and evaluate various IR models. By observing performance of your IR system using TREC\_eval result and improve the search results by tweaking the different parameters accepted by the IR model, using query parser's and boosting techniques.

## 1 Introduction

The project makes use of twitter data in three languages – English, German and Russian. Using Solr the twitter data has been indexed on 3 different cores corresponding to different IR models. We have implemented 3 following IR models:

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

For each model , the results are obtained for given queries and evaluated using relevance judgements in TREC\_eval. The output of TREC\_eval is used for comparing the models using Mean Average Precision(MAP) value.

## 2 Definitions

### 2.1 Language Model

All similarity frameworks smoothen scores based on unseen words(i.e. document length). We have used Dirichlet language model, this model does Bayesian smoothing using Dirichlet priors which implies how the actual term frequency stack up to what a Dirichlet distribution would assume to be “normal”. There is a configurable parameter ( $\mu$ ) which controls smoothing, if the value of  $\mu$  is high the scores will not change abruptly.

```
<similarity class="solr.LMDirichletSimilarityFactory">  
  <str name="mu">1000.0</str>  
</similarity>
```

## 2.2 BM25

BM25 is a type of probabilistic IR model and default model in solr, which is an upgrade of TF-IDF. There are two configurable terms: k1 and b. A higher k1 implies higher ceiling, but it also makes document length normalization more dynamic(i.e. longer documents will be penalized more). In this model, the length doesn't get multiplied to the score directly. Instead, we get ratio between document length and average length of all documents in the index. Parameter b controls length normalization : higher b makes length matter more.

```
<similarity class="solr.BM25SimilarityFactory">
  <float name="b">0.8</float>
  <float name="k1">2.0</float>
</similarity>
```

## 2.3 Divergence and Randomness Model(DFR)

DFR is a framework which includes multiple models and normalization techniques which share the same principle : term may occur in a document randomly, following a certain distribution. More the document diverges from configured random distribution, higher the score. Three components of DFR are as follows:

1. The base model, which defines random distribution.
2. An after-effect, which normalizes score of base model based on term-frequency.
3. The term frequency used by after-effect is normalized based on document length.

```
<similarity class="solr.DFRSimilarityFactory">
  <str name="c">1.0</str>
  <str name="basicModel">G</str>
  <str name="afterEffect">B</str>
  <str name="normalization">H2</str>
</similarity>
```

## 2.4 Mean Average Precision(MAP)

MAP is most standard measure among TREC community which provides a single-figure measure of quality across recall levels. While Precision is the fraction of retrieved documents that are relevant, average precision is single value obtained by averaging the precision values at each new relevant document observed.

## 3 Improving IR Model

- Advanced Query Parsing is a technique to improve search results. We have various query parsers in Solr from which we tried implementing dismax and edismax query parser. **edismax** which is Extended Dismax and improved version of Dismax query parser gave better results as it supports AND, OR, NOT operations.

- The existing Stopwords.txt For English, Russian and German was modified to remove stopwords such as “http”, “https”, “rt”.
- Used query boosters such as qf to boost the specifies fields such as text\_en^2+text\_de+text\_ru.
- We have tweaked different parameters of IR models to improve the MAP value.

1. Configurations in LM Model :

*Table 1 : Changing Parameters for LM*

Value of “mu”	MAP value
2000	0.6417
1000	0.6927

runid	all	LM
num_q	all	15
num_ret	all	279
num_rel	all	225
num_rel_ret	all	129
map	all	0.6927
gm_map	all	0.6206

*Figure 1 : MAP vale for mu=1000*

2. Configurations in BM25 Model :

*Table 2: Changing Parameters for BM25*

b	k1	MAP value
0.7	1.2	0.6843
0.5	1.5	0.6809
0.8	2.0	0.6951

runid	all	BM25
num_q	all	15
num_ret	all	279
num_rel	all	225
num_rel_ret	all	128
map	all	0.6951
gm_map	all	0.6255

Figure 2 : MAP value for  $b=0.8$  and  $k1=2.0$

3. Configurations in DFR Model : In this model, we have added a hyperparameter “c” which controls term frequency normalization with respect to document length. The default value is 1. The MAP value is 0.7018.

runid	all	DFR
num_q	all	15
num_ret	all	279
num_rel	all	225
num_rel_ret	all	129
map	all	0.7018
gm_map	all	0.6326

Figure 3: MAP value for  $c=1$

## 4 Result

The table below depicts the best MAP value for different models which is obtained after improving the IR model.

Table 3 : MAP values for different IR models

Model	MAP value
LM	0.6927
BM25	0.6951
DFR	0.7018

## 5 Conclusion

$$\text{MAP(DFR)} > \text{MAP(BM25)} > \text{MAP(LM)}$$

From the results which we obtained from various analysis it is evident that when we use DFR Model with the configured parameters and edismax query parser the result and MAP value is as expected.

## References

- [1] [https://lucene.apache.org/solr/guide/6\\_6/the-extended-dismax-query-parser.html](https://lucene.apache.org/solr/guide/6_6/the-extended-dismax-query-parser.html)
- [2] <https://sematext.com/blog/search-relevance-solr-elasticsearch-similarity/#toc-bm25-the-new-default-search-ranking-3>
- [3] [https://lucene.apache.org/solr/guide/7\\_4/other-schema-elements.html](https://lucene.apache.org/solr/guide/7_4/other-schema-elements.html)