



EVALUATION OF IR MODELS

CSE 435/535

Sravanthi Adibhatla
UBIT: sadibhat

Table of Contents

A) Implementation of BM25 in Solr 6.6.5

- I. Steps to implement with default settings
- II. Steps taken to improve performance
- III. Unsuccessful results

B) Implementation of DFR in Solr 6.6.5

- I. Steps to implement with default settings
- II. Steps taken to improve performance
- III. Unsuccessful results

C) Implementation of VSM in Solr 6.6.5

- I. Steps to implement with default settings
- II. Steps taken to improve performance
- III. Unsuccessful results

D) Results Summary

A) Implementation of BM25 model in Solr 6.6.5

I. Steps to implement with default settings:

a) Add similarity class in schema.xml

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!-- Solr managed schema - automatically generated - DO NOT EDIT -->
3 <schema name="example-data-driven-schema" version="1.6">
4   <similarity class="solr.BM25SimilarityFactory">
5     <str name="b">.75</str>
6     <str name="k1">1.2</str>
7   </similarity>
```

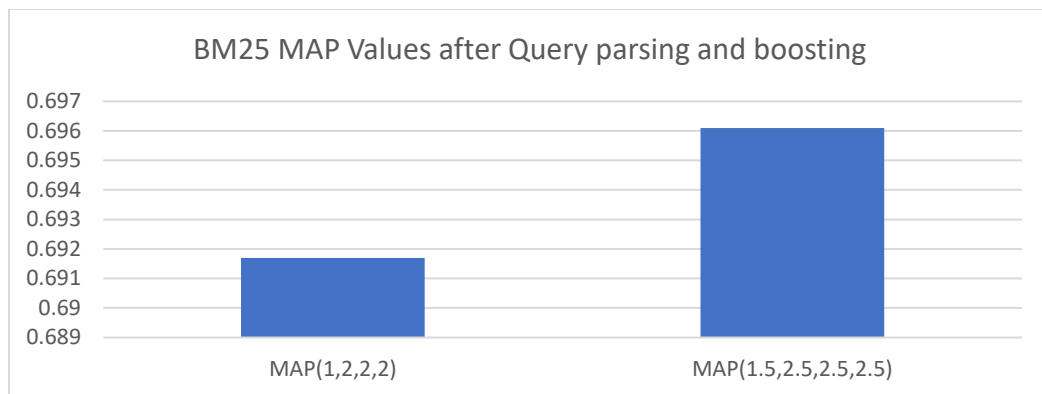
b) MAP value generated for default setting of BM25 model

```
num_q      all 15
num_ret    all 281
num_rel    all 225
num_rel_ret all 117
map        all 0.6675
gm_ap      all 0.5910
R-prec     all 0.6650
bpref      all 0.6880
recip_rank all 1.0000
ircl_prn.0.00 all 1.0000
```

II. Steps taken to improve performance:

- We did query boosting using various parameters using the query parser **Type=dismax**.
- Below are the various results for different parameters we tweaked.

BM25	Hashtag	text_en	text_ru	text_de	MAP
BM25	1	2	2	2	0.6917
BM25	1.5	2.5	2.5	2.5	0.6961



III. Unsuccessful Results:

- a) Adding synonyms has decreased the map value

No Improvement in Score

```
num_ret      all      281
num_rel      all      225
num_rel_ret  all      117
map          all      0.6675
gm_ap       all      0.5910
```

B) Implementation of DFR using Solr 6.6.5

I. Steps to implement with default settings:

a) Add similarity class in schema.xml

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <!-- Solr managed schema - automatically generated - DO NOT EDIT -->
3  <schema name="example-data-driven-schema" version="1.6">
4    <similarity class="solr.DFRSimilarityFactory">
5      <str name="basicModel">P</str>
6      <str name="afterEffect">L</str>
7      <str name="normalization">H2</str>
8      <float name="c">7</float>
9    </similarity>
```

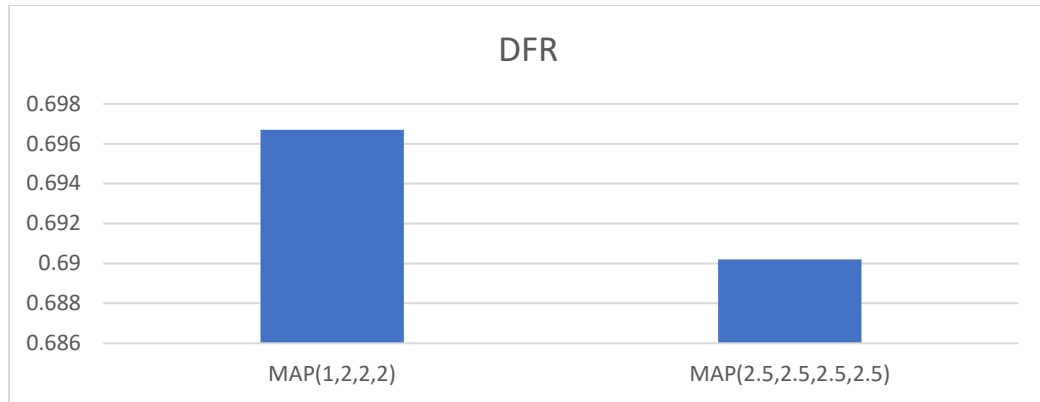
b) MAP value for DFR default

```
num_q      all      15
num_ret     all     281
num_rel     all     225
num_rel_ret all     115
map         all      0.6792
gm_ap       all      0.6008
R-prec      all      0.6920
bpref       all      0.6990
recip_rank  all      1.0000
```

II. Steps taken to improve performance:

- We did query boosting using various parameters using the query parser **Type=dismax**.
- Below are the various results for different parameters we tweaked.

DFR	Hashtag	text_en	text_ru	text_de	MAP
DFR	1	2	2	2	0.6967
DFR	2.5	2.5	2.5	2.5	0.6902



C) Implementation of VSM model in Solr 6.6.5

I. Steps to implement with default settings:

a) Add similarity class in schema.xml

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <!-- Solr managed schema - automatically generated - DO NOT EDIT -->
3  <schema name="example-data-driven-schema" version="1.6">
4    <similarity class="solr.ClassicSimilarityFactory"/>
5    <uniqueKey>id</uniqueKey>

```

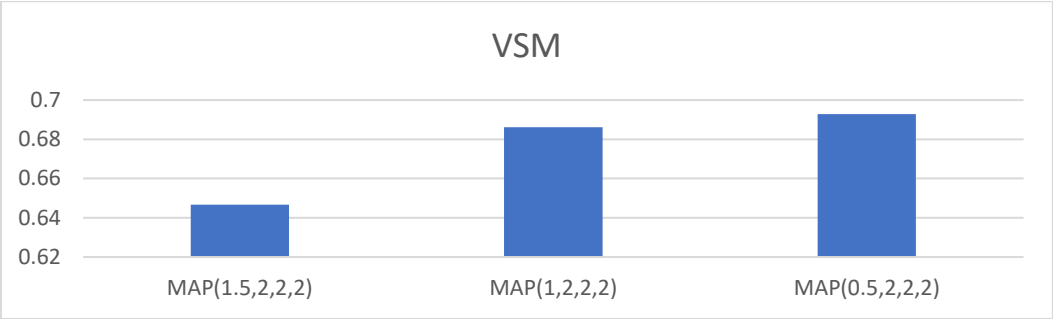
b) MAP value for VSM default

0.6639

III. Steps taken to improve performance:

- We did query boosting using various parameters using the query parser **Type=dismax**.
- Below are the various results for different parameters we tweaked.

VSM	Hashtag	text_en	text_ru	text_de	MAP
VSM	1.5	2	2	2	0.6467
VSM	1	2	2	2	0.6862
VSM	0.5	2	2	2	0.6928



Conclusion:

IR Model	Original MAP	Final MAP
DFR Model	0.6792	0.6967
BM25	0.6675	0.6961
VSM	0.6639	0.6928

