

Installation of BM25F Solr Ranking plugin

The BM25F ranking function [1] comes as a Solr Query Parser plugin [2]. In order to install it, the following steps have to be performed.

1. Create a jar of the module **./bm25f-solr-ranking** containing all the dependences
Run the command **mvn assembly:assembly** within the **./bm25f-solr-ranking** folder
will produce the jar in the target folder, (i.e. **bm25f-solr-ranking-xx-jar-with-dependencies.jar**);
2. Move the jar in the **lib** folder of the core containing the your index folder, or in the main lib of solr (see also <http://wiki.apache.org/solr/CoreAdmin>), please note that by default solr does not have the lib folder, you'll to create it.
3. Add the new query type and the new query handler, in the **solrconfig.xml** file of the core:

```
<queryParser name="bm25f"
class="eu.europeana.assets.service.ir.text.bm25f.parser.BM25FQParserPlugin">
  <!-- BY DEFAULT, QUERIES ARE PERFORMED WITH TERMS IN AND -->
  <!-- UNCOMMENT THIS IF YOU WANT QUERIES MATCHING AT LEAST THREE DIFFERENT
TERMS IN THE SAME DOCUMENT -->
    <!--
      <str name="operator">AND</str>
      <int name="minNumberOfTerms">3</int>
    -->

    <!-- UNCOMMENT THIS IF YOU WANT QUERIES PERFORMED WITH TERMS IN OR -->
    <!--
      <str name="operator">OR</str>
    -->

  <str name="mainField">text</str>
  <lst name="averageLengthFields">
    <float name="text">50</float>
    <float name="title">3</float>
    <float name="description">15</float>
    <float name="subject">9</float>
    <float name="what">8</float>
  </lst>

  <float name="k1">1.0</float>

  <lst name="fieldsBoost">
    <float name="text">3</float>
    <float name="title">39.0</float>
    <float name="description">8.0</float>
    <float name="subject">10.0</float>
    <float name="what">31.0</float>
  </lst>
  <lst name="fieldsB">
    <float name="text">0.15</float>
    <float name="title">0.05</float>
    <float name="description">0</float>
```

```
<float name="subject">0.75</float>
<float name="what">0</float>
</lst>
</queryParser>
```

The configuration file allows the admin to change the parameters.

The customizable parameters are:

- K1, the saturation factor (default 1.0)
- fieldBoost, containing the boosts to apply on the various fields;
- fieldB, containing the boosts to apply to the length of a field;
- averageLengths, the average lengths of the fields, because solr does not have this data.

Once the plugin has been plugged in, the BM25F ranking function can be called by simply adding the parameter defType=bm25f to the GET HTTP request, e.g. :

<http://mysolrmachine:8983/solr/select/?defType=bm25f&q=leonardo>

Average field lengths

For computing the scores, BM25F needs to know the average length of each field involved in the ranking. If the index changes, new average lengths can be computed using the command:

```
java -cp bm25f-solr-ranking-xx-jar-with-dependencies.jar eu.europeana.assets.service.ir.text.cli.
GetAverageFieldLengths -input solr-core-path
```

The method will iterate over the document collection and will compute the average lengths for all the fields., that you'll have to insert in the solr config file:

```
...
<lst name="averageLengthFields">
<float name="text">50</float>
<float name="title">3</float>
<float name="description">15</float>
<float name="subject">9</float>
<float name="what">8</float>
</lst> ...
```

[1] <http://nlp.uned.es/~jperezi/Lucene-BM25/>

[2] <http://wiki.apache.org/solr/SolrPlugins>