

Consuming Solr

REST API

Solr

Indexing engine, used by many large companies, netflix, stubhub etc..

Super fast indexing of documents, databases

Basic Query

Sample Public Url :

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*

Many optional parameters :

rows : equivalent of Sql 'top' or 'rownum' how many rows or documents to return
- value is number

wt : Response Format. Values of json or xml

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&wt=json&rows=2

More Basic Query

Return specific columns

fl : field list (not enabled for this specific solr instance for some reason)

indent : make it pretty, boolean

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&fl=ae_fullName&indent=true

Reference : <https://wiki.apache.org/solr/CommonQueryParameters>

Q

Query - Keyword Search / Google Like

q

Actually searches all columns/fields which are configured to be joined into a general search field (usually 'text' but not in this sample)

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?wt=json&q=AUSTRIA

FQ - Filter Query

Search a specific field.

fq

More like SQL 'where' clause (fq=fieldName:QueryText)

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&wt=json&fl=ae_officeType&fq=ae_officeType:HO

FQ cont.

Can use wildcards (*) and ranges [10 TO 20].

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&wt=json&fl=ae_officeType&fq=ae_officeType:B*

[https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&wt=json&rows=10&facet=true&fl=ae_entityEsmalD&fq=ae_entityEsmalD:\[280 TO *\]](https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&wt=json&rows=10&facet=true&fl=ae_entityEsmalD&fq=ae_entityEsmalD:[280 TO *])

Q and FQ subtleties

Solr splits words based on 'Tokenizers' (Spaces, certain punctuation etc.)

Because of this a search for North Island may come back with results of North and/or results of Island.. The configuration for the details of this is in the setup of the Solr instance.

```
<tokenizer class="solr.KeywordTokenizerFactory"/>  
<filter class="solr.StopFilterFactory" ignoreCase="true" words="stopwords.txt" />
```

Default is Case Sensitive, however, it can be changed in the configuration of the specific Solr Instance.

```
<filter class="solr.SynonymFilterFactory" synonyms="synonyms.txt" ignoreCase="true" expand="true"/>
```


Facets

Facets = Group By

facet=true

facet.field

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&wt=json&rows=1&facet=true&facet.field=ae_ncaCode

Facets continued

facet.limit : how many groupings to return (default is 100) : facet.limit=2

facet.sort : by 'count' is default, or by 'index' (sort by how they are in index)

facet.mincount : ignore items below (often used to ignore counts of 0)

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&wt=json&rows=0&facet=true&facet.field=ae_homeMemberState&facet.sort=index&facet.mincount=200

Facet Functions

Groups of Groups and functions and more

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&wt=json&fl=ae_officeType&facet=true&rows=0&facet.range=ae_entityEsmalD&facet.range.start=100&facet.range.gap=25&facet.range.end=300

Reference : https://lucene.apache.org/solr/guide/6_6/faceting.html

Reference : <http://yonik.com/solr-facet-functions/>

Stats

stats=true

stats.field

https://registers.esma.europa.eu/solr/esma_registers_mifid/select?q=*&rows=0&stats=true&stats.field=ae_entityEsmalD

Reference : https://lucene.apache.org/solr/guide/6_6/the-stats-component.html