

- solr.xml—Defines properties related to administration, logging, sharding, and SolrCloud
- solrconfig.xml—Defines the main settings for a specific Solr core
- schema.xml—Defines the structure of your index, including fields and field types

Instalando Solr

Asegurarse de la versión apropiada, en nuestro caso, java **jdk1.8.0_111** junto **jre1.8.0_111** y de **64-bits!!**

```
C:\>java -version
java version "1.8.0_111"
Java(TM) SE Runtime Environment (build 1.8.0_111-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.111-b14, mixed mode)
C:\>
```

Formas de crear un proyecto, ellas son: Standalone o bien SolrCloud...En tal caso debe darse **create_core** o bien **create_collection**

```
C:\solr-6.2.1\bin>solr.cmd start
Waiting up to 30 to see Solr running on port 8983
Started Solr server on port 8983. Happy searching!

C:\solr-6.2.1\bin>solr create -help

Usage: solr create [-c name] [-d confdir] [-n confname] [-shards #] [-replicationFactor #] [-p port]

  Create a core or collection depending on whether Solr is running in standalone
  <core> or SolrCloud
  mode <collection>. In other words, this action detects which mode Solr is running
  in, and then takes
  the appropriate action (either create_core or create_collection). For detailed
  usage instructions, do:

      bin\solr create_core -help

      or

      bin\solr create_collection -help

C:\solr-6.2.1\bin>_
```

```

C:\solr-6.2.1\bin>solr create_core -help
Usage: solr create_core [-c name] [-d confdir] [-p port]
    -c name      Name of core to create
    -d confdir   Configuration directory to copy when creating the new core, built-
in options are:
        . basic_configs: Minimal Solr configuration
        . data_driven_schema_configs: Managed schema with field-guessing support ena
bled
        . sample_techproducts_configs: Example configuration with many optional feat
ures enabled to
            demonstrate the full power of Solr
        . If not specified, default is: data_driven_schema_configs
    Alternatively, you can pass the path to your own configuration directory i
nstead of using
    one of the built-in configurations, such as: bin\solr create_core -c mycor
e -d c:/tnp/myconfig
    -p port      Port of a local Solr instance where you want to create the new cor
e
                  If not specified, the script will search the local system for a
running
                  Solr instance and will use the port of the first server it finds
.
C:\solr-6.2.1\bin>_

```

C:\solr-6.2.1>solr create_core -c de_wikipedia

Para indexar archivos xml, todos los que encuentre en la carpeta dist\..

```

C:\solr-6.2.1>java -classpath dist\solr-core-6.2.1.jar -Dauto=yes -
Dc=de_wikipedia -Ddata=files org.apache.solr.util.SimplePostTool
example\exampledocs\*.xml

```

Para indexar archivos de extension .json, .csv

```

C:\solr-6.2.1>java -classpath dist\solr-core-6.2.1.jar -Dauto=yes -
Dc=de_wikipedia -Ddata=files org.apache.solr.util.SimplePostTool
example\exampledocs\books.json (análogamente para books.csv)

```

Para indexar una gran cantidad de archivos y carpetas...en este caso todas las que se encontraron en la carpeta docs (que está en C:\solr-6.2.1) que contiene 4.399 archivos y 317 carpetas, en sólo 3 minutos y algo más !!!!.

```


C:\solr-6.2.1>java -classpath dist\solr-core-6.2.1.jar -Dauto=yes -
Dc=de_wikipedia -Ddata=files -Drecursive=yes
org.apache.solr.util.SimplePostTool docs

```

```

Posting file package-frame.html (text/html) to [base]/extract
Posting file package-summary.html (text/html) to [base]/extract
Posting file package-tree.html (text/html) to [base]/extract
Posting file package-use.html (text/html) to [base]/extract
Posting file PageTool.html (text/html) to [base]/extract
Posting file SolrParamResourceLoader.html (text/html) to [base]/extract
Posting file SolrVelocityLogger.html (text/html) to [base]/extract
Posting file SolrVelocityResourceLoader.html (text/html) to [base]/extract
Posting file VelocityResponseWriter.html (text/html) to [base]/extract
Indexing directory docs\solr-velocity\org\apache\solr\response\class-use <5 file
s, depth=6>
Posting file PageTool.html (text/html) to [base]/extract
Posting file SolrParamResourceLoader.html (text/html) to [base]/extract
Posting file SolrVelocityLogger.html (text/html) to [base]/extract
Posting file SolrVelocityResourceLoader.html (text/html) to [base]/extract
Posting file VelocityResponseWriter.html (text/html) to [base]/extract
4329 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/de_wikipedia/update.
Time spent: 0:03:02.199
C:\solr-6.2.1>_

```



- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
- de_wikipedia
 - de_wikipedia
- Dataimport
- Documents

Statistics

Last Modified: 4 days ago

Num Docs: 4376 ✓

Max Doc: 4376 ✓

Heap Memory -1

Usage:

Deleted Docs: 0

Version: 92

Segment Count: 9

Optimized: ✓

Current: ✓

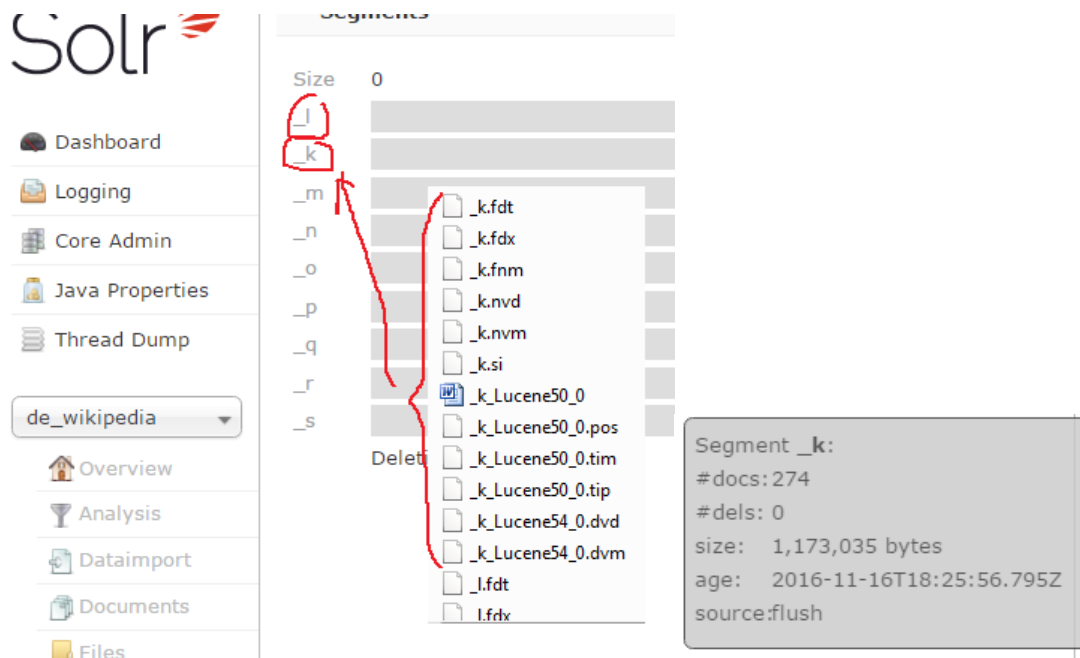
Replication (Master)

	Version
Master (Searching)	1479401
Master (Replicable)	-



Comparar con los archivos tras la indexación!!

C:\solr-6.2.1\server\solr\de_wikipedia\data\index



Vamos a AdminUI a practicar consultas sobre todos los archivos indexados..4376!!

The screenshot displays the AdminUI interface for configuring a query. On the left, a sidebar contains navigation options: Core Admin, Java Properties, Thread Dump, de_wikipedia (selected), Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query (highlighted), Replication, Schema, and Segments info. The main panel is titled '/select' and includes a 'common' tab. The query configuration fields are as follows: 'q' is set to '*:*' (circled in red); 'fq' is empty; 'sort' is empty; 'start, rows' is set to 0 and 10; 'fl' is empty; 'df' is empty; 'Raw Query Parameters' is set to 'key1=val1&key2=val2'; 'wt' is set to 'json' (circled in red); and the 'indent' checkbox is checked. To the right of the configuration, a JSON response is shown, with the 'numFound' value of 4375 circled in red. The JSON response is as follows:

```
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "q":"*:*",
      "indent":"on",
      "wt":"json",
      "_":"1479322505816"}},
  "response":{"numFound":4375,"start":0,"
    {
      "id":"C:\\solr-6.2.1\\docs\\solr-
      "stream_size":[14329],
      "date":["2016-09-15T00:00:00Z"],
      "x_parsed_by":["org.apache.tika.p
        "org.apache.tika.parser.html.Ht
      "stream_content_type":["text/html
      "dc_title":["Uses of Class org.ap
      "content_encoding":["UTF-8"],
      "content_type_hint":["text/html;
      "resourcename":["C:\\solr-6.2.1\\
      "title":["Uses of Class org.apach
      "content_type":["text/html; chars
      "_version_":1551180150687662080},
      {
        "id":"C:\\solr-6.2.1\\docs\\solr-
```

Bajo el nº de documentos encontrados

Core Admin

Java Properties

Thread Dump

de_wikipedia

Overview

Analysis

Dataimport

Documents

Files

Ping

Plugins / Stats

Query

Replication

/select

common

q

title:Use*

fq

sort

start, rows

010

fl

df

```
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "q":"title:Use*",
      "indent":"on",
      "wt":"json",
      "_":"1479322505816"},
    "response":{"numFound":1976,"start":0,"docs":[
      {
        "id":"C:\\solr-6.2.1\\docs\\solr-core\\org\\
        "stream_size":[14329],
        "date":["2016-09-15T00:00:00Z"],
        "x_parsed_by":["org.apache.tika.parser.Defa
        "org.apache.tika.parser.html.HtmlParser"]
        "stream_content_type":["text/html"],
        "dc_title":["Uses of Class org.apache.solr.
        "content encoding":["UTF-8"]1.
      }
    ]}
  }
```

Una consulta más acertiva..

Solr

Dashboard

Logging

Core Admin

Java Properties

Thread Dump

de_wikipedia

Overview

Analysis

Dataimport

Documents

Files

Ping

Plugins / Stats

Query

Replication

Schema

Segments info

q

title:*org.apache.solr.core.CloudCo
nfig*

fq

sort

start, rows

010

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

json

☒ indent

☐ debugQuery

☐ dismax

☐ edismax

☐ hl

```
"status":0,
"QTime":25,
"params":{
  "q":"title:*org.apache.solr.core.CloudConfig*",
  "indent":"on",
  "wt":"json",
  "_":"1479322505816"},
"response":{"numFound":2,"start":0,"docs":[
  {
    "id":"C:\\solr-6.2.1\\docs\\solr-core\\org\\apache\\solr\\core\\class-use\\
    "stream_size":[14329],
    "date":["2016-09-15T00:00:00Z"],
    "x_parsed_by":["org.apache.tika.parser.DefaultParser",
      "org.apache.tika.parser.html.HtmlParser"],
    "stream_content_type":["text/html"],
    "dc_title":["Uses of Class org.apache.solr.core.CloudConfig.CloudConfigBui
    "content_encoding":["UTF-8"],
    "content_type_hint":["text/html; charset=utf-8"],
    "resourcename":["C:\\solr-6.2.1\\docs\\solr-core\\org\\apache\\solr\\core\\
    "title":["Uses of Class org.apache.solr.core.CloudConfig.CloudConfigBui
    "content_type":["text/html; charset=UTF-8"],
    "_version_":1551180150687662080},
  {
    "id":"C:\\solr-6.2.1\\docs\\solr-core\\org\\apache\\solr\\core\\class-use\\
    "stream_size":[14367],
    "date":["2016-09-15T00:00:00Z"],
    "x_parsed_by":["org.apache.tika.parser.DefaultParser",
      "org.apache.tika.parser.html.HtmlParser"],
    "stream_content_type":["text/html"],
    "dc_title":["Uses of Class org.apache.solr.core.CloudConfig {Solr 6.2.1 AP:
  }
]}
```

Ante la misma consulta,

org.apache.solr.core.CloudConfig.CloudConfig

pero con Solritas, coinciden.



Find:

2 results found in 2ms Page 1 of 1	
id: C:\solr-6.2.1\docs\solr-core\org\apache\solr\core\class-use\CloudConfig.CloudConfigBuilder.html	
stream_size: 14329	
date: Thu Sep 15 00:00:00 UTC 2016	
x_parsed_by: org.apache.tika.parser.DefaultParser, org.apache.tika.parser.html.HtmlParser	
stream_content_type: text/html	
dc_title: Uses of Class org.apache.solr.core.CloudConfig.CloudConfigBuilder (Solr 6.2)	
content_encoding: UTF-8	
content_type_hint: text/html; charset=utf-8	
resourcename: C:\solr-6.2.1\docs\solr-core\org\apache\solr\core\class-use\CloudConfig.CloudConfigBuilder.html	
title: Uses of Class org.apache.solr.core.CloudConfig.CloudConfigBuilder (Solr 6.2)	
content_type: text/html; charset=UTF-8	
version: 1551180150687662080	
score: 1.0	

id: C:\solr-6.2.1\docs\solr-core\org\apache\solr\core\CloudConfig.CloudConfigBuilder.html	
stream_size: 26160	
date: Thu Sep 15 00:00:00 UTC 2016	


O en el browse, pero en formato xml

http://localhost:8983/solr/de_wikipedia/browse?&q=org.apache.solr.core.CloudConfig.CloudConfig*&wt=xml

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">5</int>
    <lst name="params">
      <str name="q">org.apache.solr.core.CloudConfig.CloudConfig*</str>
      <str name="wt">xml</str>
    </lst>
  </lst>
  <result name="response" numFound="2" start="0" maxScore="1.0">
    <doc>
      <str name="id">C:\solr-6.2.1\docs\solr-core\org\apache\solr\core\class-use\CloudConfig.CloudConfigBuilder.html</str>
      <arr name="stream_size">
        <long>14329</long>
      </arr>
      <arr name="date">
        <date>2016-09-15T00:00:00Z</date>
      </arr>
    </doc>
  </result>
</response>
```

Me fui y partí de cero, para luego darle **bin>solr start**



The image shows the Solr Admin UI. On the left is a sidebar with navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, de_wikipedia (selected), Dataimport, Documents, and Files. The main content area is divided into two sections. The top section, titled 'Statistics', displays the following information: Last Modified: about 2 hours ago, Num Docs: 4375, Max Doc: 4375, Heap Memory: -1, Usage: Deleted Docs: 0, Version: 88, Segment Count: 8, Optimized: (green checkmark), and Current: (green checkmark). The bottom section, titled 'Replication (Master)', shows a table with columns 'Version' and 'G'. The table contains two rows: 'Master (Searching)' with version '1479320797725' and 'Master (Replicable)' with a hyphen. Below the table is a green button labeled '+ Admin Extra'.

Solr

- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
- de_wikipedia
- Dataimport
- Documents
- Files

Statistics

Last Modified: about 2 hours ago
Num Docs: 4375
Max Doc: 4375
Heap Memory: -1
Usage:
Deleted Docs: 0
Version: 88
Segment Count: 8
Optimized:
Current:

Replication (Master)

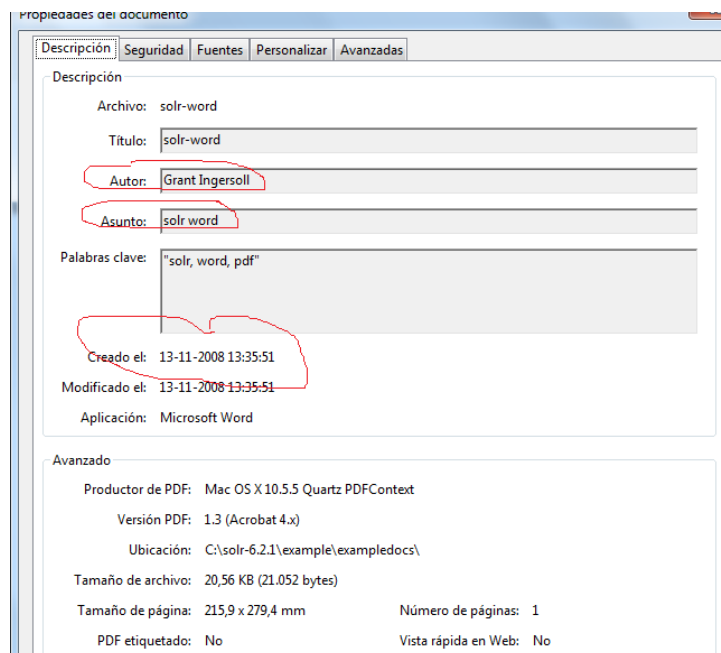
	Version	G
Master (Searching)	1479320797725	:
Master (Replicable)	-	

+ Admin Extra

Tarea pendiente, usar LINGO

<http://www.indexierung-retrieval.de/2013/02/lingo.html>

Consulta sobre el archivo PDF



The image shows a 'Propiedades del documento' (Document Properties) dialog box. It has tabs for Descripción, Seguridad, Fuentes, Personalizar, and Avanzadas. The 'Descripción' tab is active. It contains fields for Archivo: solr-word, Título: solr-word, Autor: Grant Ingersoll, Asunto: solr word, Palabras clave: "solr, word, pdf", Creado el: 13-11-2008 13:35:51, Modificado el: 13-11-2008 13:35:51, and Aplicación: Microsoft Word. The 'Avanzado' tab is also visible, showing details about the PDF production, including the producer (Mac OS X 10.5.5 Quartz PDFContext), version (1.3 Acrobat 4.x), location (C:\solr-6.2.1\example\exampledocs\), file size (20,56 KB (21.052 bytes)), page size (215,9 x 279,4 mm), number of pages (1), and whether it is PDF tagged (No) and has a fast web view (No).

Propiedades del documento

Descripción Seguridad Fuentes Personalizar Avanzadas

Descripción

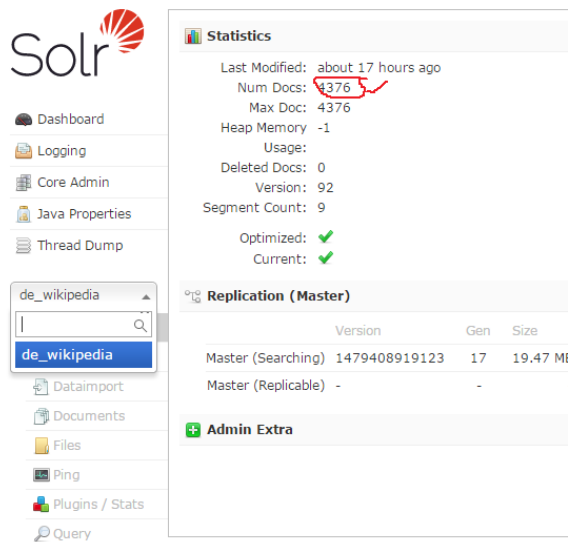
Archivo: solr-word
Título: solr-word
Autor: Grant Ingersoll
Asunto: solr word
Palabras clave: "solr, word, pdf"
Creado el: 13-11-2008 13:35:51
Modificado el: 13-11-2008 13:35:51
Aplicación: Microsoft Word

Avanzado

Productor de PDF: Mac OS X 10.5.5 Quartz PDFContext
Versión PDF: 1.3 (Acrobat 4.x)
Ubicación: C:\solr-6.2.1\example\exampledocs\
Tamaño de archivo: 20,56 KB (21.052 bytes)
Tamaño de página: 215,9 x 279,4 mm
Número de páginas: 1
PDF etiquetado: No
Vista rápida en Web: No

Indexar archivo .pdf

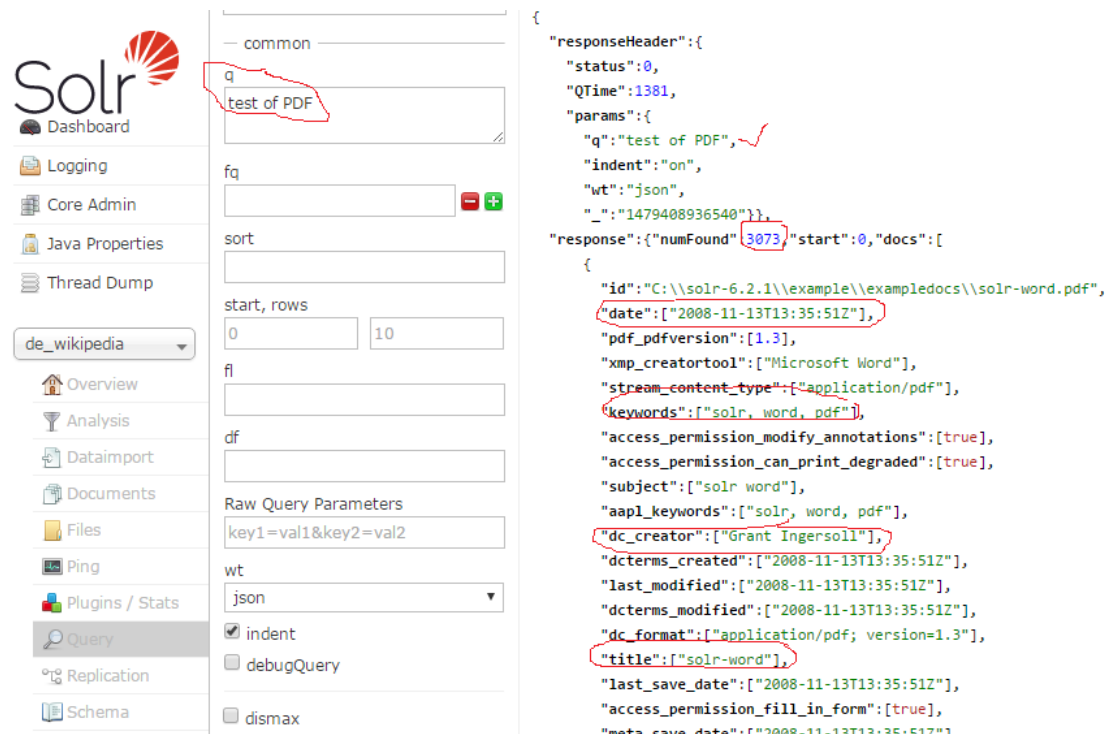
```
C:\solr-6.2.1>java -classpath dist\solr-core-6.2.1.jar -Dauto=yes -Dc=de_wikipedia -Ddata=files org.apache.solr.util.SimplePostTool example\exampledocs\*.pdf
SimplePostTool version 5.0.0
Posting files to [base] url http://localhost:8983/solr/de_wikipedia/update...
Entering auto mode. File endings considered are xml,json,jsonl,csv,pdf,doc,docx,ppt,pptx,xls,xlsx,odt,odp,ods,ott,otp,ots,rtf,htm,html,txt,log
POSTing file solr-word.pdf (application/pdf) to [base]/extract
1 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/de_wikipedia/update...
Time spent: 0:00:05.917
C:\solr-6.2.1>
```



The image shows the Solr Admin UI. On the left is a sidebar with navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, de_wikipedia (selected), Dataimport, Documents, Files, Ping, Plugins / Stats, and Query. The main content area has two tabs: 'Statistics' and 'Replication (Master)'. The 'Statistics' tab is active, showing the following data:

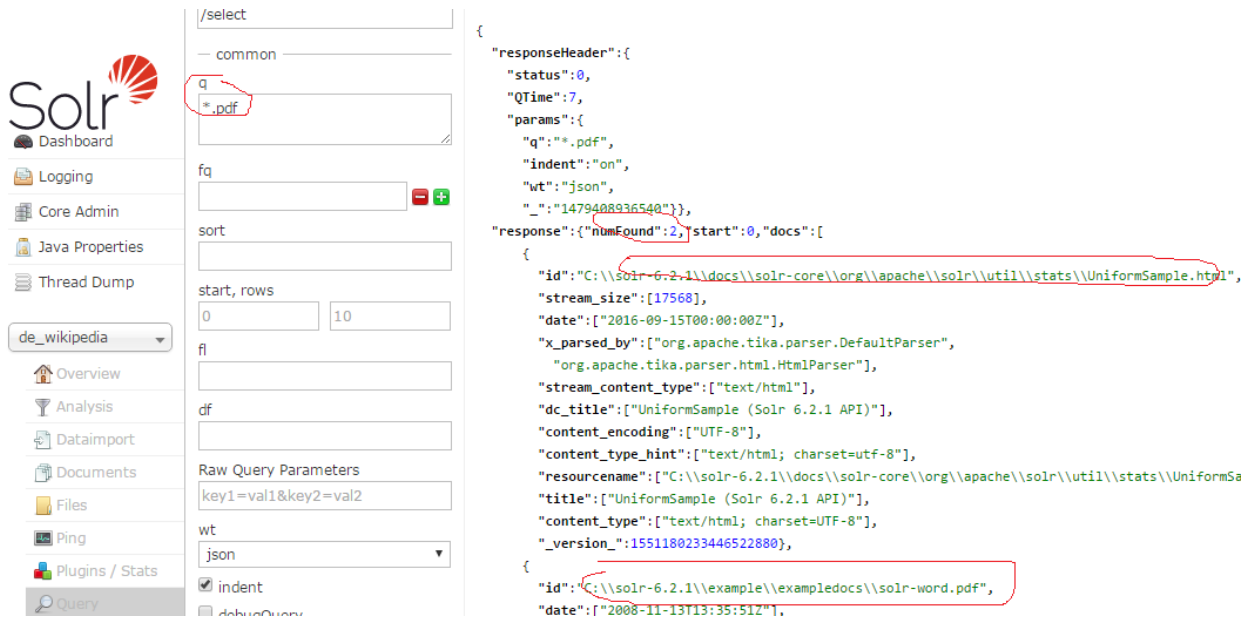
Statistics
Last Modified: about 17 hours ago
Num Docs: 4376
Max Doc: 4376
Heap Memory: -1
Usage:
Deleted Docs: 0
Version: 92
Segment Count: 9
Optimized: <input checked="" type="checkbox"/>
Current: <input checked="" type="checkbox"/>

Below the statistics is the 'Replication (Master)' tab, which shows a table with columns: Version, Gen, and Size. The table contains two rows: 'Master (Searching)' with values 1479408919123, 17, and 19.47 MB; and 'Master (Replicable)' with a dash in the Version column. Below the table is an 'Admin Extra' section.



The image shows the Solr Admin UI with the 'Query' tab selected. The sidebar is the same as in the previous image. The main content area shows the query interface. The 'q' field contains 'test of PDF'. The 'fq' field is empty. The 'sort' field is empty. The 'start, rows' fields are set to 0 and 10. The 'fl' field is empty. The 'df' field is empty. The 'Raw Query Parameters' section shows 'key1=val1&key2=val2'. The 'wt' field is set to 'json'. The 'indent' checkbox is checked. The 'debugQuery' checkbox is unchecked. The 'dismax' checkbox is unchecked. On the right side of the image is a JSON response from the query:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 1381,
    "params": {
      "q": "test of PDF",
      "indent": "on",
      "wt": "json",
      " ": "1479408936540"
    }
  },
  "response": {
    "numFound": 3073,
    "start": 0,
    "docs": [
      {
        "id": "C:\\solr-6.2.1\\example\\exampledocs\\solr-word.pdf",
        "date": ["2008-11-13T13:35:51Z"],
        "pdf_pdfversion": [1.3],
        "xmp_creatortool": ["Microsoft Word"],
        "stream_content_type": ["application/pdf"],
        "keywords": ["solr, word, pdf"],
        "access_permission_modify_annotations": [true],
        "access_permission_can_print_degraded": [true],
        "subject": ["solr word"],
        "aapl_keywords": ["solr, word, pdf"],
        "dc_creator": ["Grant Ingersoll"],
        "dcterms_created": ["2008-11-13T13:35:51Z"],
        "last_modified": ["2008-11-13T13:35:51Z"],
        "dcterms_modified": ["2008-11-13T13:35:51Z"],
        "dc_format": ["application/pdf; version=1.3"],
        "title": ["solr-word"],
        "last_save_date": ["2008-11-13T13:35:51Z"],
        "access_permission_fill_in_form": [true],
        "meta save date": ["2008-11-13T13:35:51Z"]
      }
    ]
  }
}
```



The screenshot shows the Solr Admin interface with a search query `*.pdf` entered in the `q` field. The response is a JSON object containing a `responseHeader` and a `response` object with `numFound: 2` and `start: 0`. The `docs` array contains two documents:

```

{
  "id": "C:\\solr-6.2.1\\docs\\solr-core\\org\\apache\\solr\\util\\stats\\UniformSample.html",
  "stream_size": [17568],
  "date": ["2016-09-15T00:00:00Z"],
  "x_parsed_by": ["org.apache.tika.parser.DefaultParser", "org.apache.tika.parser.html.HtmlParser"],
  "stream_content_type": ["text/html"],
  "dc_title": ["UniformSample (Solr 6.2.1 API)"],
  "content_encoding": ["UTF-8"],
  "content_type_hint": ["text/html; charset=utf-8"],
  "resourcename": ["C:\\solr-6.2.1\\docs\\solr-core\\org\\apache\\solr\\util\\stats\\UniformSa"],
  "title": ["UniformSample (Solr 6.2.1 API)"],
  "content_type": ["text/html; charset=UTF-8"],
  "_version_": 1551180233446522880,
}
{
  "id": "C:\\solr-6.2.1\\example\\exampledocs\\solr-word.pdf",
  "date": ["2008-11-13T13:35:51Z"],

```

Afinamos la consulta



The screenshot shows the Solr Admin interface with a search query `vitter.pdf` entered in the `q` field. The response is a JSON object containing a `responseHeader` and a `response` object with `numFound: 1` and `start: 0`. The `docs` array contains one document:

```

{
  "id": "C:\\solr-6.2.1\\docs\\solr-core\\org\\apache\\solr\\util\\stats\\UniformSample.html",
  "stream_size": [17568],
  "date": ["2016-09-15T00:00:00Z"],
  "x_parsed_by": ["org.apache.tika.parser.DefaultParser", "org.apache.tika.parser.html.HtmlParser"],
  "stream_content_type": ["text/html"],
  "dc_title": ["UniformSample (Solr 6.2.1 API)"],
  "content_encoding": ["UTF-8"],
  "content_type_hint": ["text/html; charset=utf-8"],
  "resourcename": ["C:\\solr-6.2.1\\docs\\solr-core\\org\\apache\\solr\\util\\stats\\UniformSample.html"],
  "title": ["UniformSample (Solr 6.2.1 API)"],
  "content_type": ["text/html; charset=UTF-8"],
  "_version_": 1551180233446522880
}

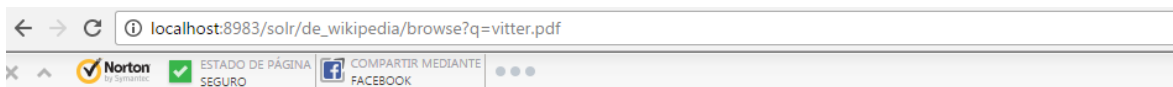
```

Annotations in the image highlight the `resourcename` and `title` fields, with text indicating that the path `C:\\solr-6.2.1\\docs\\solr-core\\org\\apache\\solr\\util\\stats\\UniformSample.html` is found in the `resourcename` field and the document `UniformSample (Solr 6.2.1 API)` is found in the `title` field.

Ahora se puede buscar por todo lo que desee utilizando la sintaxis de consultas de Solr (un superconjunto de la sintaxis de consulta de Lucene).

- Video
- *.pdf
- Vitter.pdf
- name:video
- +video +price:[* TO 400]

Ratificamos la consulta con Solritas!!



Find:

1 results found in 139ms Page 1 of 1

```
id: C:\solr-6.2.1\docs\solr-core\org\apache\solr\util\stats\UniformSample.html
stream_size: 17568
date: Thu Sep 15 00:00:00 UTC 2016
x_parsed_by: org.apache.tika.parser.DefaultParser, org.apache.tika.parser.html.HtmlParser
stream_content_type: text/html
dc_title: UniformSample (Solr 6.2.1 API)
content_encoding: UTF-8
content_type_hint: text/html; charset=utf-8
resourcename: C:\solr-6.2.1\docs\solr-core\org\apache\solr\util\stats\UniformSample.html
title: UniformSample (Solr 6.2.1 API)
content_type: text/html; charset=UTF-8
_version_: 1551180233446522880
score: 9.435654
```

En el documento PDF solr-word



Find:

1 results found in 1ms Page 1 of 1

```
id: C:\solr-6.2.1\example\exampledocs\solr-word.pdf
date: Thu Nov 13 13:35:51 UTC 2008
pdf_pdfversion: 1.3
xmp_creatortool: Microsoft Word
stream_content_type: application/pdf
keywords: solr, word, pdf
access_permission_modify_annotations: true
access_permission_can_print_degraded: true
subject: solr word
aapl_keywords: solr, word, pdf
dc_creator: Grant Ingersoll
dcterms_created: Thu Nov 13 13:35:51 UTC 2008
last_modified: Thu Nov 13 13:35:51 UTC 2008
dcterms_modified: Thu Nov 13 13:35:51 UTC 2008
dc_format: application/pdf; version=1.3
title: solr-word
```

Pero **panico** no lo acepto!!



Find:

1 results found in 2ms Page 1 of 1

id: C:\solr-6.2.1\example\exampledocs\solr-word.pdf

date: Thu Nov 13 13:35:51 UTC 2008

pdf_pdfversion: 1.3

http://localhost:8983/solr/de_wikipedia/browse?q=panico, en "formato URL"

Para ello, debí intervenir el archivo **synonyms.txt** que se encuentra en

C:\solr-6.2.1\server\solr\de_wikipedia\conf\synonyms.txt

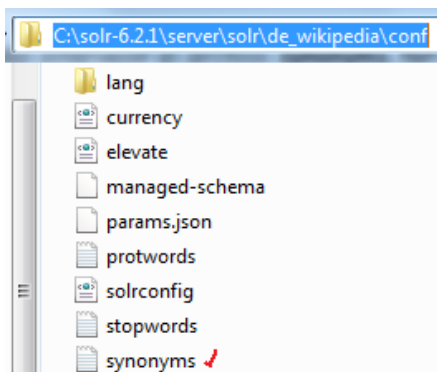
Synonym mappings can be used for spelling correction too

pixima => pixma

panico => **panic**

Debí usar **bin>solr stop -all** (bajar el servidor primero) y luego levantar el servidor con **bin>solr start**. Ahí lo captó!!

Sigamos interviniendo los archivos de la configuración



Ahora intervine **solrconfig.xml** y **schema.xml** para el caso spellchecker

Recordar que una vez alterado algún archivo de **conf** debe hacer **stop** y **start** nuevamente. Sin embargo, debo salirme completamente, esto es, matar el proceso **cmd** Windows **MS_DOS** y ahí logro llegar a **adminUI** con **de_wikipedia**. Ahora veré que resultado. Nada..grgrgrgr!!!. Para consolarme, sigo dándole a los sinónimos!!!

common

q

princesa

fq

sort

start, rows

010

fl

df

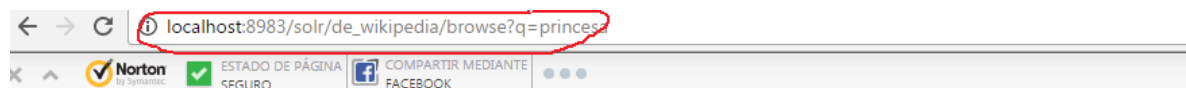
Raw Query Parameters

key1=val1&key2=val2

```
"responseHeader":{
  "status":0,
  "QTime":3,
  "params":{
    "q":"princesa",
    "indent":"on",
    "wt":"json",
    "_":"1479497057237"}},
"response":{"numFound":1,"start":0,"docs":[
  {
    "id":"0380014300",
    "cat":["book"],
    "name":["Nine Princes In Amber"],
    "price":6.99,
    "inStock":true,
    "author":["Roger Zelazny"],
    "series_t":["the Chronicles of Amber"],
    "sequence_i":1,
    "genre_s":["fantasy"],
    "_version_":1551179954209685504}}
}}
```

Synonym mappings can be used for spelling correction too
pixima => pixma
panico => panic
Princesa => princes!!!

✓ sinónimo



Find:

1 results found in 410ms Page 1 of 1

id: 0380014300

cat: book

name: Nine **Princes** In Amber

price: 6.99

inStock: true

author: Roger Zelazny

series_t: the Chronicles of Amber

sequence_i: 1

genre_s: fantasy

version: 1551179954209685504

score: 13.427476

http://localhost:8983/solr/de_wikipedia/browse?q=iPod



Find:

4 results found in 4ms Page 1 of 1

id: IW-02
name: iPod & iPod Mini USB 2.0 Cable
manu: Belkin
manu_id_s: belkin
cat: electronics, connector
features: car power adapter for iPod white



- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
- de_wikipedia
- Overview
- Analysis
- Dataimport
- Documents
- Files
- Ping
- Plugins / Stats

Request-Handler (qt)

/select

— common —

q

ipod

fq

sort

start, rows

0

10

fl

http://localhost:8983/solr/de_wikipedia/select?ind

```
{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"ipod",
      "indent":"on",
      "wt":"json",
      "_":"1479499895347"
    }
  },
  "response":{"numFound":4,"start":0,"docs":[
    {
      "id":"IW-02",
      "name":["iPod & iPod Mini USB 2.0 Cable"],
      "manu":["Belkin"],
      "manu_id_s":"belkin",
      "cat":["electronics",
        "connector"]
    }
  ]}
}
```




Logging

Core Admin

Java Properties

Thread Dump

de_wikipedia

Overview

Analysis

Dataimport

Documents

Files

Ping

Plugins / Stats

Query

Replication

Schema

☒ facet

facet.query

facet.field

cat

facet.prefix

☐ spatial

☐ spellcheck

Execute Query

```
{
  "id": "C:\\solr-6.2.1\\docs\\solr-
},
"facet_counts":{
  "facet_queries":{},
  "facet_fields":{
    "cat":[
      "book",0,
      "camera",0,
      "connector",0,
      "copier",0,
      "currency",0,
      "electronics",0,
      "electronics and computer1",0,
      "electronics and stuff2",0,
      "graphics card",0,
      "hard drive",0,
      "hardcover",0,
      "memory",0,
      "multifunction printer",0,
      "music",0,
      "paperback",0,
      "printer",0,
      "scanner",0,
      "search",0,
      "software",0]],
    "facet_ranges":{},
    "facet_intervals":{},
    "facet_heatmaps":{}}
```

☒ facet

facet.query

facet.field

cat

facet.prefix

hard

☐ spatial

☐ spellcheck

Execute Query

```
"id": "C:\\solr-6.2.
{
  "id": "C:\\solr-6.2.
},
"facet_counts":{
  "facet_queries":{},
  "facet_fields":{
    "cat":[
      "hard drive",0,
      "hardcover",0]],
    "facet_ranges":{},
    "facet_intervals":{},
    "facet_heatmaps":{}}
```

diría que hard drive fue encontrado en 1 doc

Como podemos ver en la parte superior del panel de resultados, existe un vínculo URL (indicado por la flecha) que muestra los resultados de la búsqueda en una página independiente o puede ser invocada directamente por cURL, **esto último es tarea pendiente!!**

`http://localhost:8983/solr/de_wikipedia/select?facet.field=cat&facet=on&fl=id,price,%20name,cat&fq=cat%20software&indent=on&q=price%200%20to%20400&wt=json`

Considerando los resultados, se visualizan dos **responseHeader** de nivel superior y se visualizan las secciones de **response**. El encabezado proporciona información general acerca de los detalles de la consulta y respuesta a los documentos coincidentes devueltos por Solr. Además, se muestra una faceta de búsqueda (facet_queries), en una tercera sección, donde **facet_counts**, **facet_field** son añadidos al final de los resultados devueltos. [20]

Field	Description
Request-handler (qt)	Specifies the query handler for the request. If a query handler is not specified, Solr processes the response with the standard query handler.
q	The query event. See Searching for an explanation of this parameter.
fq	The filter queries. See Common Query Parameters for more information on this parameter.
sort	Sorts the response to a query in either ascending or descending order based on the response's score or another specified characteristic.
start, rows	start is the offset into the query result starting at which documents should be returned. The default value is 0, meaning that the query should return results starting with the first document that matches. This field accepts the same syntax as the start query parameter, which is described in Searching . rows is the number of rows to return.
fl	Defines the fields to return for each document. You can explicitly list the stored fields, functions , and doc transformers you want to have returned by separating them with either a comma or a space.
wt	Specifies the Response Writer to be used to format the query response. Defaults to XML if not specified.

indent	Click this button to request that the Response Writer use indentation to make the responses more readable.
debugQuery	Click this button to augment the query response with debugging information, including "explain info" for each document returned. This debugging information is intended to be intelligible to the administrator or programmer.
dismax	Click this button to enable the Dismax query parser. See The DisMax Query Parser for further information.
edismax	Click this button to enable the Extended query parser. See The Extended DisMax Query Parser for further information.
hl	Click this button to enable highlighting in the query response. See Highlighting for more information.
facet	Enables faceting, the arrangement of search results into categories based on indexed terms. See Faceting for more information.
spatial	Click to enable using location data for use in spatial or geospatial searches. See Spatial Search for more information.
spellcheck	Click this button to enable the Spellchecker, which provides inline query suggestions based on other, similar, terms. See Spell Checking for more information.



Dashboard
Logging
Core Admin
Java Properties
Thread Dump

de_wikipedia

Overview
Analysis
Dataimport
Documents
Files
Ping
Plugins / Stats
Query
Replication
Schema

Request-Handler (qt)

/select

— common —

q

jhereg

fq

sort

start, rows

0 21

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

json

http://localhost:8983/solr/de_wikipedia/select?i

```
{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"jhereg",
      "indent":"on",
      "rows":"21",
      "wt":"json",
      "_":"1479596794967"}},
  "response":{"numFound":1,"start":0,"docs":[
    {
      "id":"0441385532",
      "cat":["book"],
      "name":["Jhereg"],
      "price":7.95,
      "inStock":false,
      "author":["Steven Brust"],
      "series_t":["Vlad Taltos"],
      "sequence_i":1,
      "genre_s":["fantasy",
      "_version_":1551179954207588352}]
  ]}
}
```

/select

— common —

q
jhereg

fq

sort

start, rows

fl

df

Raw Query Parameters
key1=val1&key2=val2

wt
json ▼

☒ indent

```
{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"jhereg",
      "spellcheck.q":"jherag",
      "indent":"on",
      "spellcheck":"on",
      "rows":"21",
      "wt":"json",
      "_":"1479596794967"}},
  "response":{"numFound":1,"start":0,"docs":[
    {
      "id":"0441385532",
      "cat":["book"],
      "name":["Jhereg"],
      "price":7.95,
      "inStock":false,
      "author":["Steven Brust"],
      "series_t":["Vlad Taltos"],
      "sequence_i":1,
      "genre_s":"fantasy",
      "_version_":1551179954207588352}]
  }}
}
```

☐ spatial
☒ spellcheck
☐ spellcheck.build
☐ spellcheck.reload

 spellcheck.dictionary

 spellcheck.count

☐ spellcheck.onlyMorePopular
☐ spellcheck.extendedResults
☐ spellcheck.collate
 spellcheck.maxCollations

 spellcheck.maxCollationTries

 spellcheck.accuracy

<!-- Spell Check

The spell check component can return a list of alternative spelling suggestions.

<http://wiki.apache.org/solr/SpellCheckComponent>

-->

<searchComponent name="spellcheck" class="solr.SpellCheckComponent">

<!-- a spellchecker built from a field of the main index -->

<lst name="spellchecker">

<str name="name">default</str>

```

<str name="field">name</str>
<str name="classname">solr.DirectSolrSpellChecker</str>

<!-- minimum accuracy needed to be considered a valid spellcheck suggestion -->
<float name="accuracy">0.5</float>

<!-- the minimum shared prefix when enumerating terms -->
<int name="minPrefix">1</int>

<!-- minimum length of a query term to be considered for correction -->
<int name="minQueryLength">3</int>
</lst>

</searchComponent>

<!-- A request handler for demonstrating the spellcheck component.

NOTE: This is purely as an example. The whole purpose of the
SpellCheckComponent is to hook it into the request handler that
handles your normal user queries so that a separate request is
not needed to get suggestions.

IN OTHER WORDS, THERE IS REALLY GOOD CHANCE THE SETUP BELOW IS
NOT WHAT YOU WANT FOR YOUR PRODUCTION SYSTEM!

See http://wiki.apache.org/solr/SpellCheckComponent for details
on the request parameters.
-->
<requestHandler name="/spell" class="solr.SearchHandler" startup="lazy">
  <lst name="defaults">

    <str name="spellcheck.dictionary">default</str>
    <str name="spellcheck">on</str>
    <str name="spellcheck.extendedResults">true</str>
    <str name="spellcheck.count">10</str>
    <str name="spellcheck.alternativeTermCount">5</str>
    <str name="spellcheck.maxResultsForSuggest">5</str>
  </lst>
</requestHandler>

```

```

</lst>
<arr name="last-components">
  <str>spellcheck</str>
</arr>
</requestHandler>

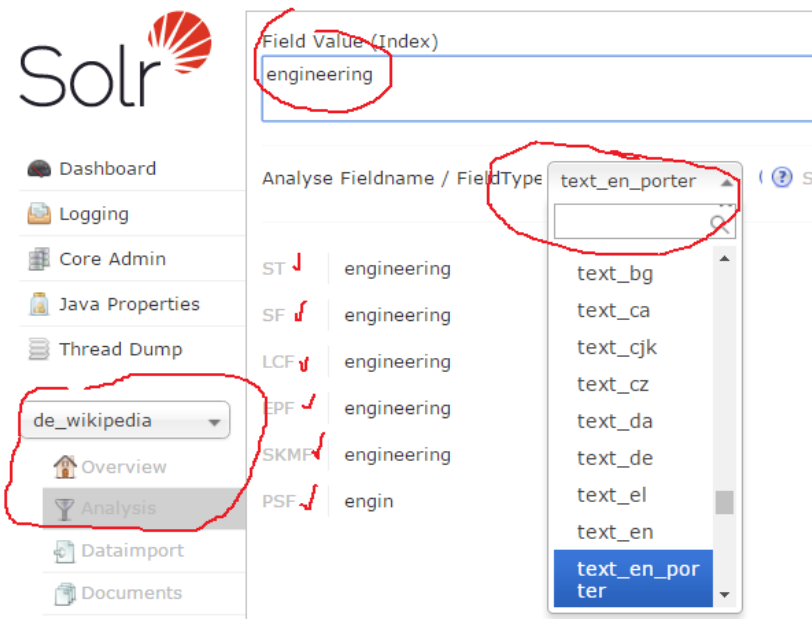
```

En el archivo **schema.xml** lo intervenimos para aplicar **porter stemming**


```

<fieldType name="text_en_porter" class="solr.TextField"
  positionIncrementGap="100">
  <analyzer>
    <tokenizer class="solr.StandardTokenizerFactory"/>
    <filter class="solr.StopFilterFactory"
      ignoreCase="true"
      words="lang/stopwords_en.txt"/>
    <filter class="solr.LowerCaseFilterFactory"/>
    <filter class="solr.EnglishPossessiveFilterFactory"/>
    <filter class="solr.KeywordMarkerFilterFactory"
      protected="protwords.txt"/>
    <filter class="solr.PorterStemFilterFactory"/>
  </analyzer>
</fieldType>

```



Notar la coincidencia!!



Solr


- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
- de_wikipedia
- Overview
- Analysis

Field Value (Index)

the schools, las escuelas

Analyse Fieldname / FieldType: text_en

ST	the	schools	las	escuelas
SF		schools	las	escuelas
LCF		schools	las	escuelas
EPF		schools	las	escuelas
SKMF		schools	las	escuelas
PSF		school	la	escuela



Solr

- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
- de_wikipedia
- Overview
- Analysis

Field Value (Index)

the schools, las escuelas

Analyse Fieldname / FieldType: text_es

ST	the	schools	las	escuelas
LCF	the	schools	las	escuelas
SF	the	schools		escuelas
SLSF	the	schools		escuel

[20] Solr UI Screen: <https://cwiki.apache.org/confluence/display/solr/Query+Screen>

Hacer correr algunos ejemplos del libro **solr in action**. Por ejemplo restaurant de cap8!!

```
C:\Program Files\Java\solr-6.2.1\bin>solr create_core -help
Usage: solr create_core [-c name] [-d confdir] [-p port]
  -c name      Name of core to create
  -d confdir   Configuration directory to copy when creating the new core, built-in options are:
    basic_configs: Minimal Solr configuration
    data_driven_schema_configs: Managed schema with field-guessing support enabled
    sample_techproducts_configs: Example configuration with many optional features enabled to demonstrate the full power of Solr
  If not specified, default is: data_driven_schema_configs
  Alternatively, you can pass the path to your own configuration directory instead of using one of the built-in configurations, such as: bin\solr create_core -c mycore -d c:/tmp/myconfig
  -p port      Port of a local Solr instance where you want to create the new core
  If not specified, the script will search the local system for a running Solr instance and will use the port of the first server it finds.
```

```
C:\Program Files\Java\solr-6.2.1\bin>solr create_core -c restaurant -d C:\Users\Eric\Desktop\Mi_Solr_general\libros_solr\solr-in-action\example-docs\ch8\cores\restaurant\conf
Copying configuration to new core instance directory:
C:\Program Files\Java\solr-6.2.1\server\solr\restaurant
Creating new core 'restaurant' using command:
http://localhost:8983/solr/admin/cores?action=CREATE&name=restaurant&instanceDir=restaurant
{
  "responseHeader":{
    "status":0,
    "QTime":4996},
  "core":"restaurant"}
```

```
C:\Program Files\Java\solr-6.2.1\bin>cd .. Indexamos, archivo restaurants.json
C:\Program Files\Java\solr-6.2.1>java -classpath dist\solr-core-6.2.1.jar -Dauto=yes -Dc=restaurant -Ddata=file org.apache.solr.util.SimplePostTool example\exampledocs\restaurants.json
SimplePostTool version 5.0.0
Posting files to [base] url http://localhost:8983/solr/restaurant/update...
Entering auto mode. File endings considered are xml,json,jsonl,cs,doc,docx,ppt,pptx,xls,xlsx,odt,odp,ods,ott,otp,ots,rtf,htm,html,txt,log
POSTing file restaurants.json (application/json) to [base]/json/docs
1 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/restaurant/update...
Time spent: 0:00:01.779
C:\Program Files\Java\solr-6.2.1>
```

common

q

,

fq

price asc

start, rows

0 20

fl

df

Raw Query Parameters

key1=val1&key2=val2

```
"responseHeader":{
  "status":0,
  "QTime":5},
"response":{"numFound":20,"start":0,"docs":[
  {
    "id":"20",
    "name":"Starbucks",
    "city":"Greenville",
    "type":"Coffee Shop",
    "state":"South Carolina",
    "tags":["coffee",
      "breakfast"],
    "price":3.0,
    "_version_":1551
  },
  {
    "id":"5",
    "name":"McDonald",
    "city":"Atlanta",
    "state":"Georgia",
    "type":"Fast Food",
    "tags":["fast food",
```

archivo restaurants.json

```
facet_fields: {
  "name": [
    "Starbucks", 6,
    "McDonalds", 5,
    "Pizza Hut", 3,
    "Red Lobster", 3,
    "Freddy's Pizza Shop", 1,
    "Sprig", 1,
    "The Iberian Pig", 1],
  "tags": [
    "breakfast", 11,
    "coffee", 11,
    "sit-down", 8,
    "fast food", 5,
    "hamburgers", 5,
    "wi-fi", 5,
    "pizza", 4,
    "delivery", 3,
    "sea food", 3,
    "gluten-free", 1,
    "pasta", 1,
    "sit down", 1,
    "southern cuisine", 1,
    "spanish", 1,
    "tapas", 1,
    "hamburguesas", 5
```

11 restaurant
tienen cafe, 5
hamburguesas, y
5 wifi

http://localhost:8983/solr/restaurants/select?q=*,*&facet=true&facet.mincount=1&facet.field=name&facet.field=tags

Facet query results showing counts for various restaurant names. The 'name' field is highlighted in red, and the text 'aparece en 6 doc' (appears in 6 docs) is written next to it.

```

{
  "_version_":1551638510749876224}}
},
"facet_counts":{
  "facet_queries":{},
  "facet_fields":{
    "name":[
      "Starbucks",6,
      "McDonalds",5,
      "Pizza Hut",3,
      "Red Lobster",3,
      "Freddy's Pizza Shop",1,
      "Sprig",1,
      "The Iberian Pig",1]],
  "facet_ranges":{},
  "facet_intervals":{},
  "facet_heatmaps":{}}
}

```

Facet query results showing counts for various restaurant names. The 'name' field is highlighted in red, and the text 'aparece en 6 doc' (appears in 6 docs) is written next to it.

Además, si hacemos facet=true y facet.field=tags y facet.prefix=s, se tiene

Facet query results showing counts for various tags. The 'tags' field is highlighted in red, and the text 'aparece en 6 doc' (appears in 6 docs) is written next to it.

```

{
  "_version_":1551638510749876224}}
},
"facet_counts":{
  "facet_queries":{},
  "facet_fields":{
    "tags":[
      "sit-down",8,
      "sea food",3,
      "sit down",1,
      "southern cuisine",1,
      "spanish",1]],
  "facet_ranges":{},
  "facet_intervals":{},
  "facet_heatmaps":{}}
}

```

Facet query results showing counts for various tags. The 'tags' field is highlighted in red, and the text 'aparece en 6 doc' (appears in 6 docs) is written next to it.

Reflexiones:

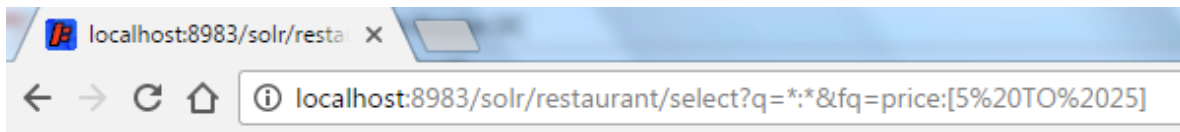
- Luego de rescatado los datos, vienen las reflexiones de los resultados, esto es, existen muchos restaurants que ofrecen "breakfast" y "coffe" que atraviesan el corpus!!
- Notar que el n° de matching con tags es mucho más que 20, pero es a consecuencia que cada documento contiene más que uno de los términos.

Table 8.1 Field faceting parameters that can be specified on the Solr URL to modify faceting behavior

Solr parameter	Possible values	Description
<code>facet</code>	<code>true</code> , <code>false</code>	Enables or disables all faceting for the current search.
<code>facet.field</code>	the name of any indexed field	Determines which field a facet should be calculated upon. This parameter may be specified multiple times to return multiple facets.
<code>facet.sort</code>	<code>index</code> , <code>count</code>	Sorts the facet values by highest number of occurrences (<code>count</code>) or by lexicographical order in the index (<code>index</code>). This parameter can be specified on a per-field basis.
<code>facet.limit</code>	An integer ≥ -1	Determines how many unique facet values will be returned for each facet. This parameter can be specified on a per-field basis.

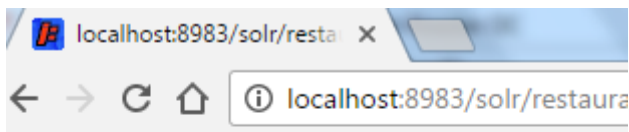
http://localhost:8983/solr/restaurant/select?q=*&facet=true&facet.mincount=1&facet.field=state&f.state.facet.limit=50&f.state.facet.sort=index&facet.field=name&f.name.facet.mincount=1&facet.field=tags&f.tags.facet.limit=5

```
{
  "facet_counts": {
    "facet_queries": {},
    "facet_fields": {
      "state": [
        "California", 4,
        "Georgia", 6,
        "Illinois", 2,
        "New York", 4,
        "South Carolina", 1,
        "Texas", 3],
      "name": [
        "Starbucks", 6,
        "McDonalds", 5,
        "Pizza Hut", 3,
        "Red Lobster", 3,
        "Freddy's Pizza Shop", 1,
        "Sprig", 1,
        "The Iberian Pig", 1],
      "tags": [
        "breakfast", 11,
        "coffee", 11,
        "sit-down", 8,
        "fast food", 5,
        "hamburgers", 5]],
    "facet_ranges": {},
    "facet_intervals": {},
    "facet_histograms": {}
  }
}
```



```
{
  "responseHeader":{
    "status":0,
    "QTime":232},
  "response":{"numFound":11,"start":0,"docs":[
    {
      "id":"2",
      "name":"Red Lobster",
      "city":"Atlanta, GA",
      "type":"Sit-down Chain",
      "state":"Georgia",
      "tags":["sea food",
        "sit-down"],
      "price":22.0,
      "_version_":1551638510714224640},
    {
      "id":"4",
      "name":"McDonalds",
      "city":"San Francisco, CA",
      "type":"Fast Food"
```

http://localhost:8983/solr/restaurants/select?q=*&fq=price:[5 TO 25]&facet=true&
facet.query=state: ("New York" OR "Georgia" OR "South Carolina") &
facet.query=state: ("Illinois" OR "Texas") &
facet.query=state: ("California")

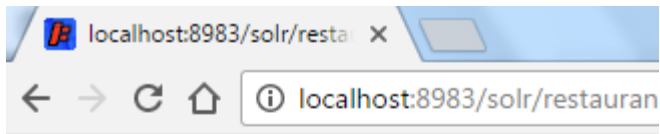


```
{
  "responseHeader":{
    "status":0,
    "QTime":419},
  "response":{"numFound":11,"start":0,"docs":[
    {
      "id":"2",
      "name":"Red Lobster",
      "city":"Atlanta, GA",
      "type":"Sit-down Chain",
      "state":"Georgia",
      "tags":["sea food",
        "sit-down"],
      "price":22.0,
      "_version_":1551638510714224640},
    {
      "id":"4",
      "name":"McDonalds",
      "city":"San Francisco, CA",
      "type":"Fast Food",
      "state":"California",
```

```

http://localhost:8983/solr/restaurants/select?q=*&rows=0&facet=true&
facet.query=price:[* TO 5]&
facet.query=price:[5 TO 10]&
facet.query=price:[10 TO 20]&facet.query=price:[20 TO 50]&
facet.query=price:[50 TO *]

```



```

{
  "responseHeader":{
    "status":0,
    "QTime":31},
  "response":{"numFound":20,"start":0,"docs":[]
},
  "facet_counts":{
    "facet_queries":{
      "price:[* TO 5]":6,
      "price:[5 TO 10]":5,
      "price:[10 TO 20]":3,
      "price:[20 TO 50]":6,
      "price:[50 TO *]":0},
    "facet_fields":{},
    "facet_ranges":{},
    "facet_intervals":{},
    "facet_heatmaps":{}}
}

```

No estamos listos para sugerir en base a los precios.???, o de acuerdo a la distancia??

```

http://localhost:8983/solr/restaurants/select?q=*&facet=true&
facet.field=state&
facet.field=city&
facet.query=price:[* TO 10]&
facet.query=price:[10 TO 25]&
facet.query=price:[25 TO 50]&
facet.query=price:[50 TO *]

```

```

''
"facet_counts":{
  "facet_queries":{
    "price:[* TO 10]":11,
    "price:[10 TO 25]":5,
    "price:[25 TO 50]":4,
    "price:[50 TO *]":0},
  "facet_fields":{
    "state":[
      "Georgia",6,
      "California",4,
      "New York",4,
      "Texas",3,
      "Illinois",2,
      "South Carolina",1],
    "city":[
      "Atlanta, GA",6,
      "New York, NY",4,
      "Austin, TX",3,
      "San Francisco, CA",3,
      "Chicago, IL",2,
      "Greenville, SC",1,
      "Los Angeles, CA",1]},
  "facet_ranges":{},
  "facet_intervals":{},
  "facet_heatmaps":{}}

```

Tf-idf???