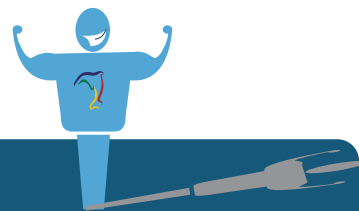


¿Qué ofrece Autentia Real Business Solutions S.L?

Somos su empresa de **Soporte a Desarrollo Informático**.
 Ese apoyo que siempre quiso tener...

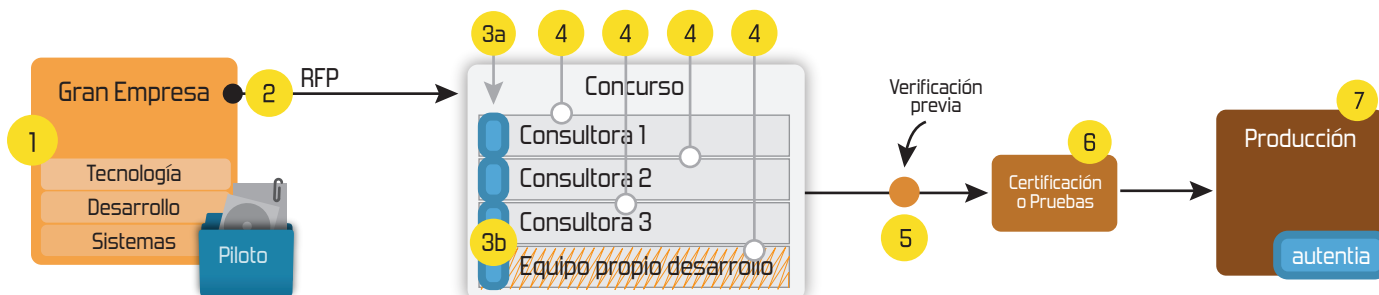
1. Desarrollo de componentes y proyectos a medida



2. Auditoría de código y recomendaciones de mejora

3. Arranque de proyectos basados en nuevas tecnologías

1. Definición de frameworks corporativos.
2. Transferencia de conocimiento de nuevas arquitecturas.
3. Soporte al arranque de proyectos.
4. Auditoría preventiva periódica de calidad.
5. Revisión previa a la certificación de proyectos.
6. Extensión de capacidad de equipos de calidad.
7. Identificación de problemas en producción.



4. Cursos de formación (impartidos por desarrolladores en activo)

Spring MVC, JSF-PrimeFaces /RichFaces,
 HTML5, CSS3, JavaScript-jQuery

Gestor portales (Liferay)
 Gestor de contenidos (Alfresco)
 Aplicaciones híbridas

Tareas programadas (Quartz)
 Gestor documental (Alfresco)
 Inversión de control (Spring)

Control de autenticación y
 acceso (Spring Security)
 UDDI
 Web Services
 Rest Services
 Social SSO
 SSO (Cas)

JPA-Hibernate, MyBatis
 Motor de búsqueda empresarial (Solr)
 ETL (Talend)

Dirección de Proyectos Informáticos.
 Metodologías ágiles
 Patrones de diseño
 TDD

BPM (jBPM o Bonita)
 Generación de informes (JasperReport)
 ESB (Open ESB)

Estás en:

[Inicio](#) [Tutoriales](#) [Introducción a Apache Solr.](#)



DESARROLLADO POR:

[Jose Manuel Sánchez Suárez](#)

Consultor tecnológico de desarrollo de proyectos informáticos.

Puedes encontrarme en [Autentia](#): Ofrecemos servicios de soporte a desarrollo, factoría y formación

Somos expertos en Java/J2EE

[Catálogo de servicios Autentia](#)



Fecha de publicación del tutorial: 2009-02-26



[Share](#) |

[Regístrate para votar](#)

Introducción a Apache Solr.

0. Índice de contenidos.

- 1. Introducción.
- 2. Entorno.
- 3. Primeros pasos.
- 4. Indexación y recuperación de documentos.
- 5. Referencias.
- 6. Conclusiones.

1. Introducción

Apache Solr es un motor de búsqueda de código abierto que proporciona una capa de abstracción sobre Apache Lucene. Solr se define como la "serverización" de Lucene.

Sus características son las siguientes:

- permite realizar peticiones HTTP para indexar o consultar documentos: se podría decir que tiene un api estilo REST, aunque no hace uso de todos los verbos, sí permite la recuperación de documentos en formato XML y JSON,
- incluye caches internas para devolver con mayor rapidez el resultado de las consultas,
- incluye una administración web que permite: consultar estadísticas de rendimiento, incluyendo el uso de cache, realizar búsquedas mediante un formulario, navegar por los términos más populares del índice, visualizar un desglose detallado de las matemáticas de puntuación y las fases de análisis de texto. Si habéis trabajado antes con Lucene se asemeja bastante a la aplicación [Luke](#).
- permite la configuración de la indexación y recuperación de documentos mediante ficheros de configuración xml: añade una librería de analizadores textuales a los que provee por defecto Lucene, introduce el concepto de campo tipado, lo que permite introducir fechas y mejorar la ordenación,
- añade mejoras a las consultas básicas de Lucene,
- incluye navegación por facetas en las búsquedas,

Últimas Noticias

[theEvt 2011: Evento de Tecnología & Negocio en la Web](#)

[Proxima charla en TheEvt: La Technicenta, de programador a empresario](#)

[XVI Charla Autentia - Refactoring y Clean Code - Cambio de fecha](#)

[XV Charla Autentia - web2py \(y Google App Engine\) - Vídeos y Material](#)

[XVI Charla Autentia - Refactoring y Clean Code](#)

[Histórico de NOTICIAS](#)

Últimos Tutoriales

[Session Timeout en JSF2 con el soporte de Primefaces.](#)

[Cambiando el plugin de eclipse para](#)

- integra el plugin de Lucene para la recuperación de contenido con resaltado de sintaxis en las coincidencias,
- dispone de un plugin de "spell check" o revisión gramatical, para realizar recomendaciones de búsqueda,
- dispone de un plugin de [búsqueda de documentos similares](#),
- permite el manejo de documentos ricos (word, pdf, ...) basándose en el proyecto Apache Tika, basado en el antiguo [Lius](#),
- esta escrito en Java, se trata de una aplicación web que se puede desplegar en cualquier contenedor de servlets,
- esta preparado para su despliegue en alta disponibilidad.

Recientemente han liberado la versión 3.1, haciendo coincidir el número de versión con la última versión de Lucene y entre sus nuevas características podemos encontrar:

- soporta geolocalización en los documentos, de modo que podemos realizar búsquedas con un filtro de distancia,
- soporta como formato de respuesta CSV,
- podemos restringir la navegación de facetas por rangos numéricos,
- incorpora un componente de autocompletado que denominan "suggester",
- permite la ordenación de resultados haciendo uso de funciones,
- incorpora una web de ejemplo basada en plantillas de velocity que permite acceder a todas estas nuevas funcionalidades <http://localhost:8983/solr/admin/>,

En este tutorial vamos a mostrar cómo descargar y hacer correr la última versión de Apache Solr y cómo realizar una prueba de concepto indexando y recuperando sus documentos de prueba a través de su propia interfaz de búsqueda.

2. Entorno.

El tutorial está escrito usando el siguiente entorno:

- Hardware: Portátil MacBook Pro 17' (2.93 GHz Intel Core 2 Duo, 4GB DDR3 SDRAM).
- Sistema Operativo: Mac OS X Snow Leopard 10.6.7
- Apache Solr 3.1.

3. Primeros pasos.

Lo primero es descargarnos Solr de la siguiente ubicación

<http://www.apache.org/dyn/closer.cgi/lucene/solr/>; nos descargamos un fichero comprimido y lo descomprimos en disco.

La distribución de Solr viene con un servidor jetty que nos permite arrancar una versión con un mínimo esfuerzo. Otra cosa será como instalar Solr en producción, pero eso se escapa del objetivo de este tutorial.

Dentro de la distribución de Solr podemos encontrar un directorio example con un Solr instalado en un jetty, para arrancarlo basta con ejecutar **java -jar start.jar** en dicho directorio. Si todo va bien tendremos una salida por consola con esta última línea:

```
1 | 2011-04-25 11:57:29.218:INFO::Started SocketConnector@0.0.0.0:8983
```

Indicándonos el puerto para acceder al jetty que contiene la aplicación web de Solr.

Para acceder a la administración de Solr, basta con acceder a la siguiente url:
<http://localhost:8983/solr/admin/>

Maven, de IAM a m2eclipse.

 Primeros pasos con github: subir un proyecto al repositorio.

 Desarrollo de aplicaciones mixtas (web/nativa) en Android

 Habilitar autocompletado de etiquetas para JSF en un proyecto Eclipse gestionado por Maven.

Últimos Tutoriales del Autor

 Session TimeOut en JSF2 con el soporte de Primefaces.

 Primeros pasos con github: subir un proyecto al repositorio.

 Habilitar autocompletado de etiquetas para JSF en un proyecto Eclipse gestionado por Maven.

 Listener del ciclo de vida de JSF, en JSF2 y con el soporte de FacesTrace (Primefaces)

 Log4J: Modificar el nivel de log en tiempo de ejecución con el soporte de Spring

Síguenos a través de:



Últimas ofertas de empleo

2011-03-02
 T. Información - Analista / Programador - MALAGA.

2011-02-24
 T. Información - Especialista CRM - MADRID.

2011-02-16
 Marketing - Experto en Marketing - CADIZ.

2011-02-08
 Comercial - Ventas - CADIZ.

2011-01-28

Solr Admin (example)

192.168.1.14:8983
 cwd=/Applications/dev/apache-solr-3.1.0/example SolrHome=solr/
 HTTP caching is OFF

Solr [SCHEMA] [CONFIG] [ANALYSIS] [SCHEMA BROWSER]
 [STATISTICS] [INFO] [DISTRIBUTION] [PING] [LOGGING]

App server: [JAVA PROPERTIES] [THREAD DUMP]

Make a Query [FULL INTERFACE]

Query String:

Assistance [DOCUMENTATION] [ISSUE TRACKER] [SEND EMAIL]
 [SOLR QUERY SYNTAX]

Current Time: Sat Apr 23 01:13:52 CEST 2011
 Server Start At: Sat Apr 23 00:57:28 CEST 2011

Para acceder a la nueva página de búsqueda de Solr, basta con acceder a la siguiente url:
<http://localhost:8983/solr/browse/>

[Solr Admin](#)



Examples: Simple [Spatial](#)

Find:

☐ Boost by Price

Field Facets

cat
manu_exact

0 results found in 9
ms Page 0 of 0

0 results found. Page
0 of 0

Query Facets

Range Facets

Price (in \$)

[More Than 600.0](#) (0)

Manufacture Date

Options: [enable debug](#) [enable annotation XML](#)
 Generated by [VelocityResponseWriter](#)
 Documentation: [Solr Home Page](#), [Solr Wiki](#)
 Disclaimer: The locations displayed in this
 demonstration are purely fictional. It is more than
 likely that no store with the items listed actually
 exists at that location!

El índice está vacío, con lo que lo primero es indexar documentos.

4. Indexación y recuperación de documentos.

Comercial - Ventas -
SEVILLA.



Jose Manuel
Sánchez
[sanchezsuares](#)

Problemas con IAM y
el famoso updating
classpath infinito ???
Cámbiate a m2eclipse
-->
<http://tinyurl.com/6d3l>
yesterday · reply · retweet ·
favorite

@adictosaltrabaj |
<http://t.co/CZan2Pz>
> Session Timeout en
#JSF2 con el soporte
de #primefaces.
yesterday · reply · retweet ·
favorite

@adictosaltrabaj |
<http://ow.ly/4HCZ6> >
Primeros pasos con
github: subir un
proyecto al



Join the conversation

Dentro del directorio `example/exampledocs` de la distribución de Solr podemos encontrar documentos para indexar y realizar pruebas. Para indexarlos todos basta con ejecutar el script `./post.sh *.xml` que se conecta al api REST de Solr y realiza un update del documento xml que contiene cada uno de los ficheros.

```
01 <add><doc>
02   <field name="id">MA147LL/A</field>
03   <field name="name">Apple 60 GB iPod with Video Playback Black</field>
04   <field name="manu">Apple Computer Inc.</field>
05   <field name="cat">electronics</field>
06   <field name="cat">music</field>
07   <field name="features">iTunes, Podcasts, Audiobooks</field>
08   <field name="features">Stores up to 15,000 songs, 25,000 photos, or 150
09   hours of video</field>
10   <field name="features">2.5-inch, 320x240 color TFT LCD display with LED
11   backlight</field>
12   <field name="features">Up to 20 hours of battery life</field>
13   <field name="features">Plays AAC, MP3, WAV, AIFF, Audible, Apple Lossless,
14   H.264 video</field>
15   <field name="features">Notes, Calendar, Phone book, Hold button, Date
16   display, Photo wallet, Built-in games, JPEG photo playback, Upgradeable
17   firmware, USB 2.0 compatibility, Playback speed control, Rechargeable
18   capability, Battery level indication</field>
19   <field name="includes">earbud headphones, USB cable</field>
20   <field name="weight">5.5</field>
21   <field name="price">399.00</field>
22   <field name="popularity">10</field>
23   <field name="inStock">true</field>
24   <!-- Dodge City store -->
25   <field name="store">37.7752,-100.0232</field>
26   <field name="manufacturedate_dt">2005-10-12T08:00:00Z</field>
27 </doc></add>
```

La estructura de documento xml hace referencia a nombres de campos y contenido de los mismos.

En la instalación de Solr de ejemplo vienen preconfigurados una serie de campos para nuestros documentos, comunes para todos, con una tipología. Dichos campos los podemos encontrar en el documento xml de configuración **example/solr/conf/schema.xml**

Los campos susceptibles de formar parte de los documentos indexados se basan en esa definición de tipos de datos. Las características de los campos son las comunes en Lucene: indexed, stored, name,... pero además permite campos multivaluados, permite la definición dinámica de campos, permite la copia del contenido de un campo en otro, para facilitar el uso de distintos tipos de filtros textuales....

La definición de los analizadores o filtros se realiza en Solr a nivel de definición del tipo de dato. Permite declarar una cadena de analizadores, los que provee de base Lucene y añade algunos otros.

Con el comando anterior lo que hemos hecho es insertar una serie de productos en Solr, podríamos insertar más, en formato xml y, como estamos en versión 3.1, también en formato JSON.

Para recuperarlos, vamos a acceder primero a la opción de búsqueda de la página de administración de Solr. Podemos buscar por **ipod** obteniendo, por defecto, este tipo de resultado.

```
- <response>
- <lst name="responseHeader">
  <int name="status">0</int>
  <int name="QTime">12</int>
  - <lst name="params">
    <str name="indent">on</str>
    <str name="start">0</str>
    <str name="q">ipod</str>
    <str name="version">2.2</str>
    <str name="rows">10</str>
  </lst>
</lst>
- <result name="response" numFound="3" start="0">
  - <doc>
    - <arr name="cat">
      <str>electronics</str>
      <str>connector</str>
    </arr>
```

La url que consultamos es esta: [http://localhost:8983/solr/select/?q=ipod&version=2.2&start=0&rows=10&indent=on](http://localhost:8983/solr/select?q=ipod&version=2.2&start=0&rows=10&indent=on)

Aunque también podemos cambiar el tipo de respuesta a JSON, añadiendo a la url el parámetro wt=json

```

{
  "responseHeader":{
    "status":0,
    "QTime":2,
    "params":{
      "indent":"on",
      "start":0,
      "q":"ipod",
      "wt":"json",
      "version":"2.2",
      "rows":"10"}},
  "response":{"numFound":3,"start":0,"docs":[
    {
      "id":"IW-02",
      "name":"iPod & iPod Mini USB 2.0 Cable",
      "manu":"Belkin",
      "weight":2.0,
      "price":11.5,
      "popularity":1,
      "inStock":false,
      "store":"37.7752,-122.4232",
      "manufacturedate_dt":"2006-02-14T23:55:59Z",
      "cat":["electronics","connector"],
      "features":[
        "car power adapter for iPod, white"]},
    {
      "id":"F8V7067-APL-KIT",
      "name":"Belkin Mobile Power Cord for iPod w/ Dock",
      "manu":"Belkin",
      "weight":4.0,
      "price":19.95,
      "popularity":1,
      "inStock":false,
      "store":"45.17614,-93.87341",
      "manufacturedate_dt":"2005-08-01T16:30:25Z",
      "cat":["electronics","connector"],
      "features":[
        "car power adapter, white"]},
    {
      "id":"MA147LL/A",
      "name":"Apple 60 GB iPod with Video Playback Black",
      "manu":"Apple Computer Inc.",
      "includes":"earbud headphones, USB cable",
      "weight":5.5,
      "price":399.0,
      "popularity":10,
      "inStock":true,
      "store":"37.7752,-100.0232",
      "manufacturedate_dt":"2005-10-12T08:00:00Z",
      "cat":["electronics","music"],
      "features":[
        "iTunes, Podcasts, Audiobooks",
        "Stores up to 15,000 songs, 25,000 photos, or 150 hours of video",
        "2.5-inch, 320x240 color TFT LCD display with LED backlight",
        "Up to 20 hours of battery life",
        "Plays AAC, MP3, WAV, AIFF, Audible, Apple Lossless, H.264 video",
        "Notes, Calendar, Phone book, Hold button, Date display, Photo wallet,
    ]}
  ]}
}

```

La url que consultamos es esta otra: [http://localhost:8983/solr/select/?q=ipod&version=2.2&start=0&rows=10&indent=on&wt=json](http://localhost:8983/solr/select?q=ipod&version=2.2&start=0&rows=10&indent=on&wt=json)

Si probamos con wt=csv, el formato de respuesta será un CSV.

Si accedemos a la página de búsqueda decorada y facetada buscando por el mismo término, obtendremos los mismos resultados con la siguiente información adicional y el siguiente formato:

Find:

[Enviar consulta](#)

[Restablecer](#)

☒ Boost by Price

Field Facets

cat

[Electronics](#) (3)

[Connector](#) (2)

[Music](#) (1)

manu_exact

[Belkin](#) (2)

[Apple Computer Inc.](#) (1)

Query Facets

[Ipod](#) (3)

[GB](#) (1)

Range Facets

Price (in \$)

[0.0](#) (2)

[350.0](#) (1)

[More Than 600.0](#) (0)

Manufacture Date

[Ene 2005](#) (2)

[Ene 2006](#) (1)

3 results found in 30 ms Page 1 of 1

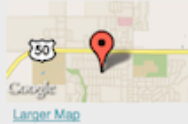
Apple 60 GB iPod with Video Playback Black [More](#)

[Like This](#)

Price: 399,00 €

Features: iTunes, Podcasts, Audiobooks Stores up to 15,000 songs, 25,000 photos, or 150 hours of video 2.5-inch, 320x240 color TFT LCD display with LED backlight Up to 20 hours of battery life Plays AAC, MP3, WAV, AIFF, Audible, Apple Lossless, H.264 video Notes, Calendar, Phone book, Hold button, Date display, Photo wallet, Built-in games, JPEG photo playback, Upgradeable firmware, USB 2.0 compatibility, Playback speed control, Rechargeable capability, Battery level indication

In Stock: true



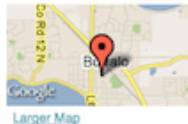
Belkin Mobile Power Cord for iPod w/ Dock [More](#)

[Like This](#)

Price: 19,95 €

Features: car power adapter, white

In Stock: false



iPod & iPod Mini USB 2.0 Cable [More Like This](#)

Price: 11,50 €

Features: car power adapter for iPod, white



A resaltar:

- el listado de productos decorado con una miniatura generada haciendo uso del api de google maps para geolocalizar la ubicación de la tienda física,
- el resultado facetado por la categoría de producto y la manufactura, esto es, el fabricante,
- un ejemplo de facetado por rango de precio y de fecha de fabricación

5. Referencias.

- <http://lucene.apache.org/solr/>

6. Conclusiones.

La potencia de Lucene accesible como un servicio en red mediante un protocolo HTTP ligero, ¿qué más se puede pedir?, que soporte distintos analizadores para un mismo campo a través de copias, que soporte distintos analizadores para indexación y búsqueda, que soporte búsqueda por facetas y además por rangos, que soporte búsqueda por documentos similares, que soporte geolocalización y búsqueda con filtro de distancia... lo dicho ¿qué más se puede pedir?

En breve veremos cómo indexar y recuperar documentos haciendo uso del api de java. Stay tuned!

Un saludo.

Jose

jmsanchez@autentia.com

Anímate y coméntanos lo que pienses sobre este **TUTORIAL**:

Puedes opinar o comentar cualquier sugerencia que quieras comunicarnos sobre este tutorial; con tu ayuda, podemos ofrecerte un mejor servicio.

[Enviar comentario](#)

(Sólo para usuarios registrados)

» **Regístrate** y accede a esta y otras ventajas «

COMENTARIOS



Esta obra está licenciada bajo [licencia Creative Commons de Reconocimiento-No comercial-Sin obras derivadas 2.5](#)

Copyright 2003-2011 © All Rights Reserved | [Texto legal y condiciones de uso](#) | [Banners](#) | [Powered by Autentia](#) | [Contacto](#)

