

Apache Solr Search

Examines the benefits of using an external search engine, taking search out of the database. Setup, configuration and special features

Audrey Foo, [Raincity Studios](#)
Vancouver Drupal Users Group
Jan 2009

What is Apache Solr?

Solr is an open source enterprise search server based on the Lucene Java search library and extends it.

- * with XML/HTTP and JSON APIs
- * hit highlighting
- * faceted search
- * caching
- * replication
- * a web administration interface
- * optimized for high volume web traffic

Schema.xml

Configuration of fields to store, index, facet with schema.xml, more intelligent processing,

Schema organized into 3 sections:

types

example types: string, text, sint, boolean

defined and related to the underlying Solr class

fields

example fields: url, title, keywords, creationDate, content

dynamic fields no need to pre-define every field

fields options by use case

other declarations

unique key

Sample Solr add command

```
<add>
```

```
  <doc>
```

```
    <field name="url">http://localhost/myBlog/solr-rocks.html</field>
```

```
    <field name="title">Solr Search is Simply Great</field>
```

```
    <field name="keywords">solr,lucene,enterprise,search</field>
```

```
    <field name="creationDate">2007-01-06T05:04:00.000Z</field>
```

```
    <field name="rating">10</field>
```

```
    <field name="content">Solr is a really great open source search server. It scales,  
it's easy to configure and the Solr community is really supportive.</field>
```

```
    <field name="published">on</field>
```

```
  </doc>
```

```
</add>
```

Apache solr install with Jetty

- * [download solr](#) and extract
- * cd {apache solr directory}/example/
- * \$ java -jar start.jar (assumes java 1.5)
- * scroll to <http://localhost:8983/solr/admin>
- * install [apache solr module](#) and enable the module
- * copy the schema.xml to {apache solr directory}/example/solr/conf
- * wipe search index (admin/settings/search/wipe and admin/settings/apachesolr/index)
- * restart jetty
- * re-run cron.php

Tomcat and Solr

- * Tomcat vs jetty useful for multiple sites
- * Requirements: Java 1.5 or greater, tomcat 5.5.x
- * Better to start from scratch rather than with a package.
- * Init script for starting, stopping solr with tomcat
- * copy the war file, and make a solr home for each webapp
- * Easy setup without editing:
tomcat home/solr/conf/bin
- * On ubuntu, use aptitude and move around dpkg

Apachesolr **Term Modifiers**

Wildcards

* for multiple character wildcard

? for single character wildcard

Fuzzy Searches

~ at the end of a single word

~0.8 value between 0 and 1, 1 with higher similarity

Proximity Searches

"jakarta apache"~10 search for apache and jakarta within 10 words of each other

Range searches

mod_date:[20020101 to 20030101]

title:{Aida To Carmen}

Boosting a Term (from 0 upwards, default is 1)

jakarta^4 apache

"jakarta apache"^4 "apache lucene"

Faceting

- * grouping of data by a certain field, showing an overview of counts.
- * apache solr module, provides a handy block,
- * faceting is done on indexed rather than sorted values, no need to store.
- * are not tokenized into separate words

Getting search results from apachesolr e.g. for autocomplete

If using database for autocomplete, may use sql with 'abc%', with apachesolr search use 'abc*'

When using apachesolr, to search a keyword from all nodes of type profile

```
$results = apachesolr_search_search('search',  
  strtolower($text) .* 'type:profile');
```

Loop through \$results for each search result which contains an array of information such as title, snippet, nid

Custom data indexed

* **Apachesolr Organic Groups Integration**

- * `hook_apachesolr_update_index(&$document, $node)`
- * `hook_apachesolr_og_facets()`
- * `hook_block()`
- * `hook_apachesolr_process_results()`
- * `hook_theme()`

Security

Solr does not concern itself with authenticating requests, the authentication should either be enforced by Tomcat, or by some middleware.

* use Tomcat's basic authentication scheme.

References

Djun Kim - great resource on many topics

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

<http://wiki.apache.org/jakarta-lucene/FrontPage>

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

<http://www.us.apachecon.com/presentation/materials/18/apache-sc>

<http://www.ibm.com/developerworks/java/library/j-solr1/>

<http://www.ibm.com/developerworks/web/library/wa-lucene2/>

<http://acquia.com/blog/drupals-search-framework-the-execution-a-s>

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

instructions for localsolr setup