MergingSolrIndexes

Merging Solr Indexes

Sometimes you have more than one Solr index and you want to merge them together into a single index.

Merging Through CoreAdmin

⚠️ Solr1.4

CoreAdminHandler now supports merging one or more indexes into another index (since Solr 1.4).

```
http://localhost:8983/solr/admin/cores?action=mergeindexes&core=core0&indexDir=/opt/solr/core1/data
/index&indexDir=/opt/solr/core2/data/index
```

The above command will merge the indexes of core1 and core2 into core0. The path for this command is the 'adminPath' defined in solr.xml (default is /admin/cores).

Before executing this command, one must make sure to call commit on core1 and core2 (in order to close IndexWriter) and no writes should happen on core1 and core2 until the merge command completes. Failure to do so may corrupt the core0 index. Once the merge is completed, a commit should be called on core0 to make the changes visible to searchers.

NOTE: In this example core0 must exist and have a compatible schema with core1 and core2. The 'mergeindexes' command will not create a new core named 'core0' if it does not already exist.

⚠️ Solr3.3 CoreAdminHandler also supports merging one or more cores into another core (since Solr 3.3) through a "srcCore" parameter.

```
http://localhost:8983/solr/admin/cores?action=mergeindexes&core=core0&srcCore=core1&srcCore=core2
```

The differences between using "srcCore" parameter and "indexDir" parameter are that:

1. Using the "indexDir" parameter one can merge indexes not associated with a Solr Core e.g. indexes built directly via Lucene
2. Using the "indexDir" parameter, one must take care that the index is not being written to - this means closing IndexWriter or if it is a solr core's index, issuing a commit command.
3. The "indexDir" must be the path to an index directory on the disk of the solr host - this makes it cumbersome. On the other hand, one can just give the core name with "srcCore" parameter
4. Using "srcCore", care is taken to ensure that the merged index is not corrupted even if writes are happening in parallel on the source index

Merging Through Lucene IndexMergeTool

Another way is to use the IndexMergeTool that comes as part of lucene-misc. In order to do this:

1. Find the lucene jar file that your version of solr is using. You can do this by copying your solr.war file somewhere and unpacking it (jar xvf solr. war). Your lucene jar file should be in WEB-INF/lib. It is probably called something like lucene-core-2007-05-20_00-04-53.jar. Copy it somewhere easy to find.
2. Download a copy of lucene from http://www.apache.org/dyn/closer.cgi/lucene/java/ and unpack it. The file you're interested in is contrib/misc /lucene-misc-VERSION.jar
3. Make sure both indexes you want to merge are closed.
4. Issue this command: `java -cp /path/to/lucene-core-VERSION.jar:/path/to/lucene-misc-VERSION.jar org.apache. lucene.misc.IndexMergeTool /path/to/newindex /path/to/index1 /path/to/index2` This will create a new index at `/path/to /newindex` that contains both index1 and index2. Copy this new directory to the location of your application's solr index (move the old one aside first, of course) and start solr.

Example

The command below assumes that the files `lucene-core-3.4.0.jar` and `lucene-misc-3.4.0.jar` are in the current directory:

```
java -cp lucene-core-3.4.0.jar:lucene-misc-3.4.0.jar org.apache.lucene.misc.IndexMergeTool ./newindex ./app1
/solr/data/index ./app2/solr/data/index
```

Caution:
Lucene MergeTool and Solr CoreAdmin both will give result in an index with duplicate docs as a result of merge; if there are 2 docs with the same uniqueKey in 2 shards to be merged.