SimpleTextCodecExample

Setting up SimpleTextCodec

⚠️ Solr4.0+ only. New to 4.0 is the ability to create per-field codecs. An example of this is the SimpleTextCodec that is distributed with the solr **source** code. However, the codecs aren't part of the binary distribution, which has caused some confusion. These instructions will allow you to use the SimpleText Codec as an exemplar.

Setup

You'll need Apache Ant and Apache Subversion. NOTE: your machine may already have these installed, try

```
svn -h
ant -h
```

if you get the help output, you're good to go.

Steps:

- Get the source code
- Build the example
- Build the codec jar
- Modify the solronfig.xml file
- Modify your schema.xml file

Get the source code

Just follow the instructions at: How To Contribute. The short form is to execute the following comand:

```
svn checkout http://svn.apache.org/repos/asf/lucene/dev/trunk
```

for trunk, or:

```
svn checkout http://svn.apache.org/repos/asf/lucene/dev/branches/branch_4x
```

for the 4.x branch.

We'll call the directory all this got checked out into SOLR_CODE which will probably be something like <where you checked things out>/branch_4x

Build the example

Now you need to build the example code. Note: this produces the same code as is present in the "example" directory in the Solr distro.

```
cd SOLR_CODE/solr
ant example
```

This may take a while. You may be prompted to execute a separate step to install Apache Ivy if you don't already have it on your computer. If you don't, the instructions to install it will be printed out on the screen when you type "ant example". Follow them and re-execute "ant example".

You should see "BUILD SUCCESSFUL" eventually.

Build the codec jar

Here's where it gets a bit tricky. The SimpleTextCodec is **not** built by the step above. So here's what you do:

```
cd SOLR_CODE/lucene/codecs
ant
```

Again, you should see "BUILD SUCCESSFUL" printed out. But just above that you should see:
"Building jar: SOLR_CODE/lucene/build/codecs/lucene-codecs-<version>.jar". This is the jar file that you'll need to have , make a note of it.

Modify the solronfig.xml file

This file is located in SOLR_CODE/solr/example/solr/collection1/conf. There are a couple of things you need to do

- Make the jar available to Solr next time you start it.
- Load the CodecFactory when Solr starts

Make the jar available to Solr next time you start it

Add a line like this. I put this after the other <lib> directives, but it's pretty arbitrary as long as it's a direct child of <config>.

```
<lib dir="../../../../lucene/build/codecs/" />
```

Load the CodecFactory when Solr starts

Add a line like this. Again where this goes is arbitrary, it just has to be a direct child of <config>. This causes Solr to load this class at startup.

```
<codecFactory name="CodecFactory" class="solr.SchemaCodecFactory" />
```

Modify your schema.xml file

This file is located in SOLR_CODE/solr/example/solr/collection1/conf

Whew! all that is preliminary. The rest is more straight-forward. You have to define a fieldType that uses the coded and you have to use that fieldType in some of your fields. NOTE: it is NOT necessary to use these in *all* your fields, you can specify codecs on a per-field basis.

You only have two more steps...

- Add a new fieldType using the SimpleTextCodec
- Use the new fieldType in some fields

Add a new fieldType using the SimpleTextCodec

Add something like this to the <types> section

```
<fieldType name="string_simpletext" class="solr.StrField" postingsFormat="SimpleText" />
```

This is not a very interesting fieldType, notice it's based on the "StrField" which means that it's not analyzed in any way, so searching is only for the exact input. Of course you can use fieldTypes with analysis chains like this. Note that this is based on TextField.

```
<fieldType name="text_simpletext" class="solr.TextField" postingsFormat="SimpleText">
  <analyzer>
    <tokenizer class="solr.WhitespaceTokenizerFactory"/>
    <filter class="solr.LowerCaseFilterFactory"/>
  </analyzer>
</fieldType>
```

Use the new fieldType in some fields

Add some lines like this to the <fields> section

```
<field name="simple_string" type="string_simpletext" indexed="true" stored="true"/>
<field name="simple_text" type="text_simpletext" indexed="true" stored="true"/>
```

At this point, you should have the SimpleText results in your SOLR_CODE/solr/example/solr/collection1/data/index directory, look for files of the form: *SimpleTest*.pst

As always, the first time someone actually follows instructions deficiencies pop out. Feel free to modify this page with whatever clarifications you think would be helpful.