

SHRINE 4.1.0 Chapter 10.3 - Generate SHRINE Lucene Search and Autosuggest Indices

These instructions will show how to generate Lucene search and auto suggest indices from the tables in the i2b2 **i2b2metadata** database. Typically one admin - usually the hub admin - will take responsibility for generating the indices and share them with all other nodes in the network.

Download the Lucene indexing tool which is distributed as an executable jar file from here:

<https://repo.open.catalyst.harvard.edu/nexus/content/groups/public/net/shrine/shrine-lucene-indexer/4.1.0/shrine-lucene-indexer-4.1.0-jar-with-dependencies.jar>

Create the input files:

1. Export the data in the **TABLE_ACCESS** table and also the data from all the tables listed in the *c_table_name* column of the **TABLE_ACCESS** table. Any delimiter can be used in the exported files but the default used in the indexing tool is '|'. There should be a header line in the exported data files. A " (quotation mark) should be used as the escape character. Place the files for each exported table into a directory.
 - a. The exported **TABLE_ACCESS** is required to have the following header columns:
 - i. c_table_cd
 - ii. c_table_name
 - iii. c_hlevel
 - iv. c_fullname
 - v. c_name
 - vi. c_synonym_cd
 - vii. c_visualattributes
 - viii. c_basecode
 - ix. c_metadataxml
 - x. c_tooltip
 - a. The other exported ontology tables are required to have the following header columns:
 - i. c_hlevel
 - ii. c_fullname
 - iii. c_name
 - iv. c_synonym_cd
 - v. c_visualattributes
 - vi. c_basecode
 - vii. c_metadataxml
 - viii. c_tooltip
 - ix. m_applied_path
2. Create a category definition file that defines the categories for the concepts in the ontology. The category definition file containing in the following order the concept path (i.e. a concatenation of the C_TABLE_CD and C_FULLNAME columns from **TABLE_ACCESS**), code category(optional), concept category, and the code set(optional). The separator for this file can be specified with the **-r** flag but by default the separator is a comma. The code category is the category that the concept is displayed under the **Medical Concept List**. The concept category is the label displayed next to the concept in the query definition panel. The code set is the category that the concept appears in when filtering search results. There should be no header line in the category definition file.
 - a. Here is an example category definition file.

Category Definition File

```
\\i2b2_DEMO\\i2b2\\Demographics, Demographics, Demographic
\\i2b2_DIAG\\i2b2\\Diagnoses, Diagnoses, Diagnosis, Diagnoses ICD9
\\i2b2_EXPR\\i2b2\\Expression Profiles Data, Expression Profiles Data
\\i2b2_LABS\\i2b2\\Labtests, Labs, Laboratory Test
\\i2b2_MEDS\\i2b2\\Medications, Medications, Medication
\\i2b2_PROC\\i2b2\\Procedures, Procedures, Procedure
\\i2b2_PROV\\i2b2\\Providers, Providers, Provider
\\i2b2_REP\\i2b2\\Reports, Reports, Report
\\i2b2_VISIT\\i2b2\\Visit Details, Visit Details, Visit Detail
\\ICD10_ICD9\\Diagnoses, Diagnoses, Diagnosis, Diagnoses ICD10
\\PHI\\PHI, PHI, Protected Health Information
\\BIRN\\BIRN, BIRN, BIRN
\\Custom Metadata\\Custom Metadata, Custom Metadata
```

Create the Lucene search index :

```
java -jar shrine-ontology-lucene-indexer-4.1.0-jar-with-dependencies.jar -o <insert exported data directory> -c <insert category definition file>
```

Create the Lucene autosuggest index :

```
java -jar shrine-ontology-lucene-indexer-4.1.0-jar-with-dependencies.jar -a -o <insert exported data directory> -c <insert category definition file>
```

Additional information

Specify a different delimiter for the exported ontology files. In the following example a tab delimiter is specified instead:

```
java -jar shrine-ontology-lucene-indexer-4.1.0-jar-with-dependencies.jar -p '\t' -o <insert exported data directory> -c <insert category definition file>
```

List all available command line options :

```
java -jar shrine-ontology-lucene-indexer-4.1.0-jar-with-dependencies.jar
```

In Result, the search bar should now be able to autosuggest your input

begin typing criteria

report|

reports 9 concepts

diagnosis reported 4 concepts

diagnosis reported without 4 concepts

reported 4 concepts

reported without 4 concepts

Reports