Creating New Structure Hasher and Standardizer in GSRS 3.x

Danny Katzel  
January 2022

Note: the code used for this example is located in the NCATS internal project:

<https://github.com/ncats/gsrs3-lychi-extension/>

What is Structure Hasher and Structure Standardizer?

Stand-alone extension jar

Create new Maven project (does not need to be spring boot just a normal java maven project)

The Substance Structure Standardizer and Hasher interfaces are in the gsrs-modules-substances-core so you will need to add that as a dependency in your maven pom. Since this jar will be put inside another GSRS 3 instance you can make this a compile time only scope since the substances-core classes will be present wherever you are using this new code.

We will make a Structure Hasher based on NCATS’ Lychi tool:

Here are the relevant parts of the new project’s Maven pom.xml file:

<**groupId**>gov.nih.ncats</**groupId**>  
<**artifactId**>gsrs3-lychi-extension</**artifactId**>  
<**version**>3.0-SNAPSHOT</**version**>

<**properties**>  
 <**gsrs.substances.version**>3.0-SNAPSHOT</**gsrs.substances.version**>  
</**properties**>

<**dependency**>  
 <**groupId**>gov.nih.ncats</**groupId**>  
 <**artifactId**>gsrs-module-substances-core</**artifactId**>  
 <**version**>${gsrs.substances.version}</**version**>  
 <**scope**>compile</**scope**>  
</**dependency**>

Then write your classes:

|  |
| --- |
| **public class** LychiStructureHasher **implements** StructureHasher {   @Override  **public void** hash(Chemical chem, String mol, BiConsumer<String, String> keyValueConsumer) {  … } |

Once your code is written you can use maven to package up the jar file. This jar will be very small, only the few classes you write.

You can use mvn install to add the jar to your local maven repository or add it to a build script that takes the built jar and runs the mvn install-file on it so that another system that builds and deploys your GSRS system can use it. Or finally you may choose to push your artifact to the Maven Central Repository so it’s accessible anywhere around the world.

Either way, in your GSRS 3 instance you will be able to use it by specifying it as a new dependency

<**dependency**>  
 <**groupId**>gov.nih.ncats</**groupId**>  
 <**artifactId**>gsrs3-lychi-extension</**artifactId**>  
 <**version**>3.0-SNAPSHOT</**version**>  
</**dependency**>

Finally, in your GSRS conf file you can tell GSRS to use these new standardizer and hasher implementations:

**ix.structure-hasher** = **"ix.core.chem.LychiStructureHasher"  
ix.structure-standardizer** = **"ix.core.chem.LychiStandardizer"**