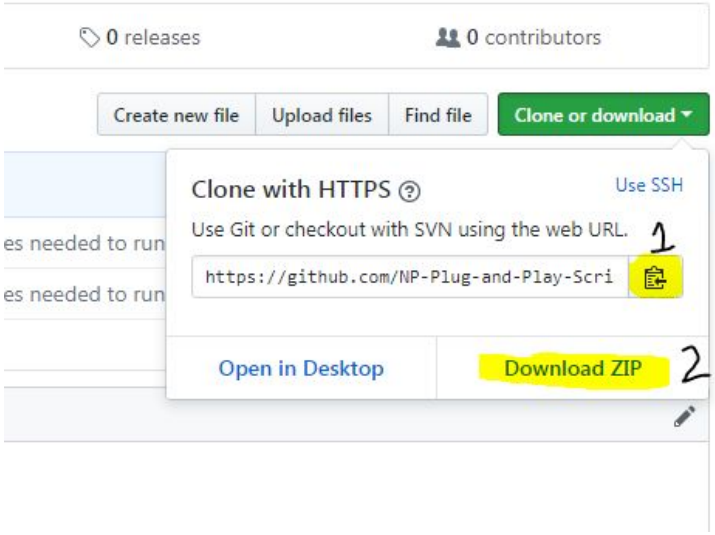
Natural Product Database Extractor

Introduction

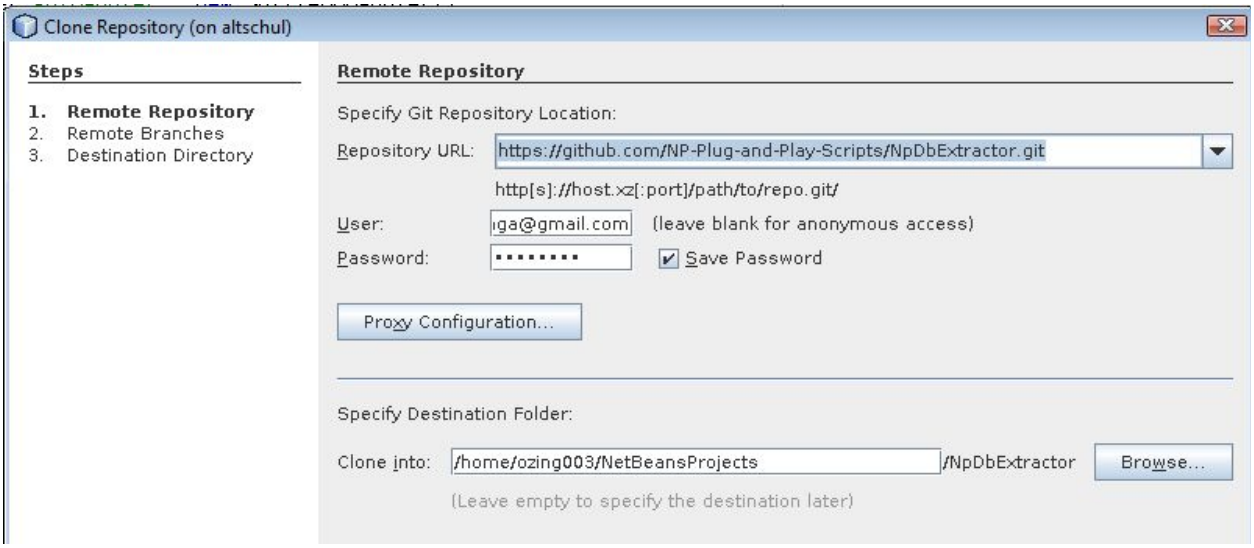
In a perfect world one would never have to see inside a sql database, this tool makes this a reality for the NP Database. By using a gui created with Java the program connects to a database of choice and after selecting the (super/sub) classes you want you can then search to get all the entries for those set values. The results are then displayed in a table on the screen where the results can be reviewed. If the user is satisfied with the results it can then be saved. This gives the user the results in a csv file, which can then be used for further research. Currently the tool only gives you the ID and the SMILES string back in the csv this can be changed in the code however. The tool was primarily made to be used for the cfm-id pipeline which requires a csv input containing the ID and SMILES string.

Installation

Before you start installation make sure you have the correct NP Database (or change the code so it works for your DB as explained below). For the tool to work straight out of the the github. The database should contain a table named structures with at least the following columns; structure_id, smile, superclass,class,subclass. Any additional information is fine as long as these 5 column are in the db.



Next get the repository from https://github.com/NP-Plug-and-Play-Scripts/NpDbExtractor use a Java IDE such as Netbeans or Eclipse (netbeans is installed on altschul) to create a repository at your favorite location. Do this by in netbeans clicking on Team > Git >Clone (or Remote > Clone) and then select the Specify Git Repository Location option. Enter the URL obtained in the Repository URL field.



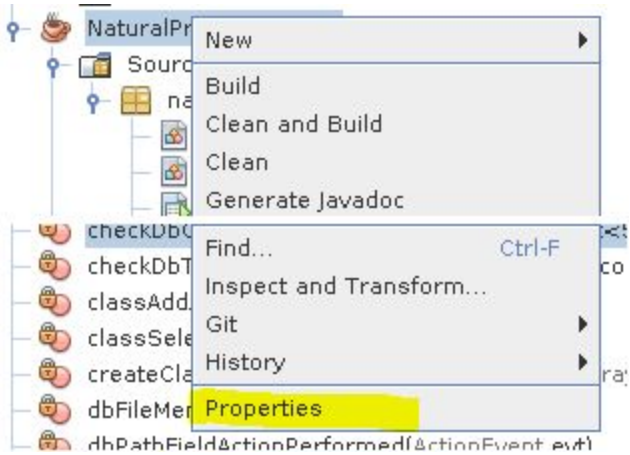
Once entered select a location to clone it to in the “Clone into” field. Press “Finish” and a clone will now be made to the location of preference.

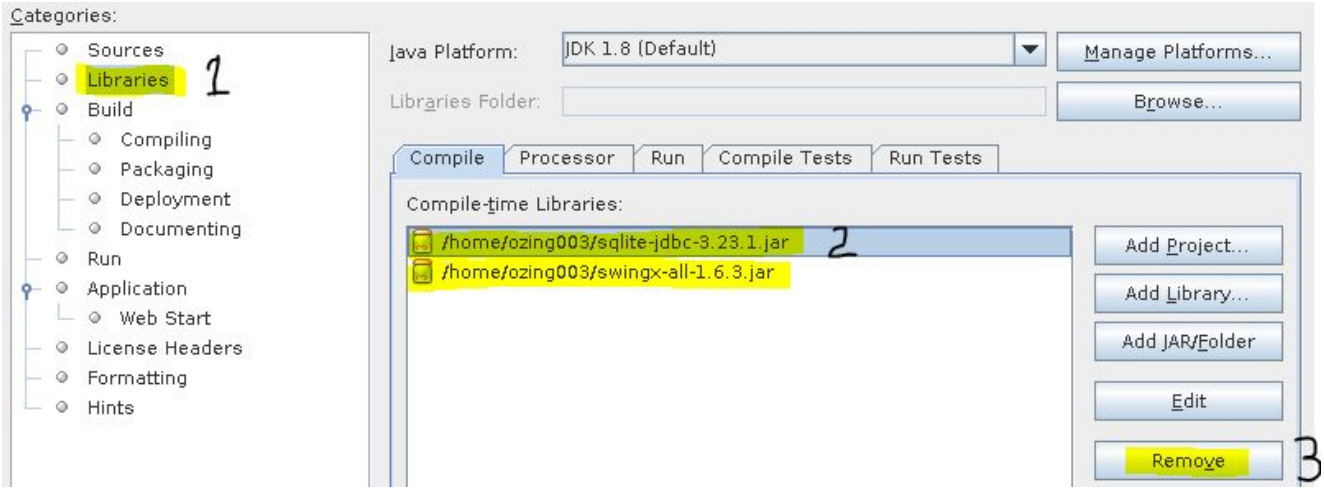
Next open the new repository and you will notice that most likely it has 2 errors. These are because the libraries are missing.
To make the project run you will need to download these 2 libraries.

- sqlite-jdbc-3.23.1.jar https://bitbucket.org/xerial/sqlite-jdbc/downloads/ newer versions should work too.
- Swingx-all-1.6.3.jar http://www.java2s.com/Code/Jar/s/Downloadswingxall163jar.htm newer versions might work too.

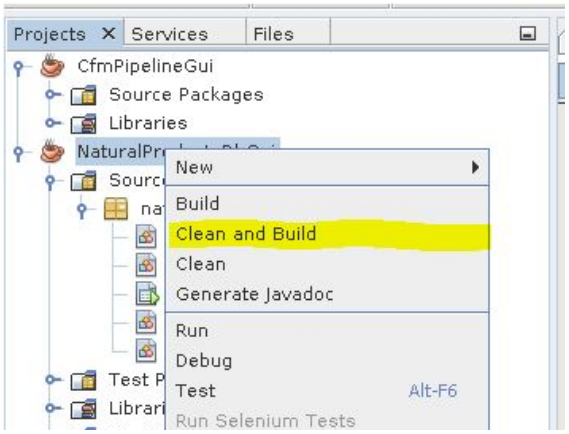
Right click the project and click properties. Go to the library tab (1) select the libraries(2) in there and remove (3) them both.

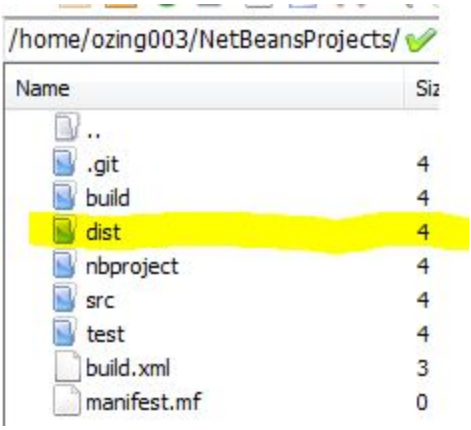
Next add the libraries you downloaded with the “Add JAR/Folder” option. Go to the path you downloaded the .Jar files and add both of them. Click ok and the issues should be resolved.





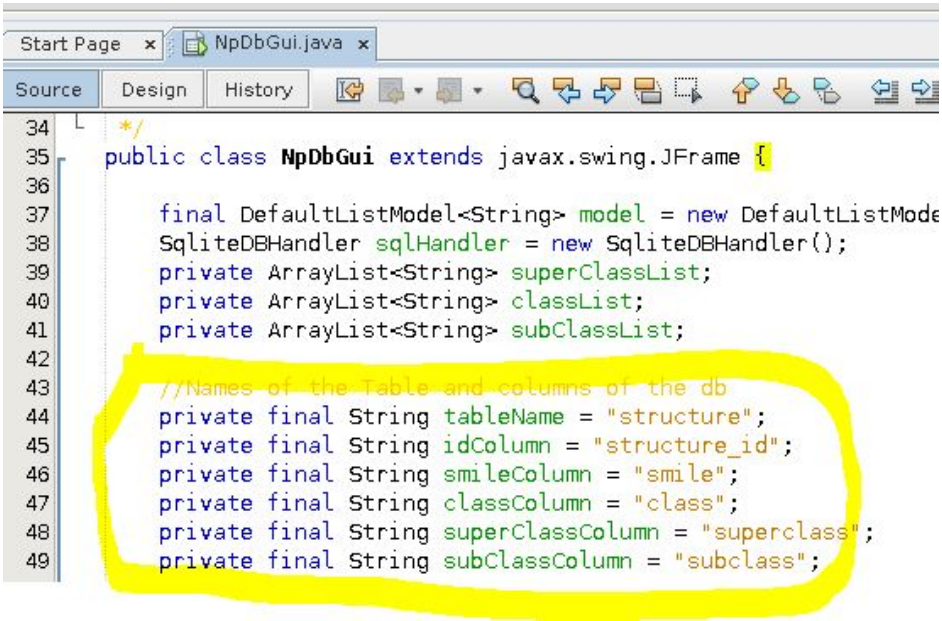
Next build the project using the IDE by right clicking the project and select clean and build. This way a directory “dist” will be made in the Netbeans folder on your pc. Another way get the runnable jar is by downloading the zip from this link: https://github.com/NP-Plug-and-Play-Scripts/NpDbExtractor-runnable and extract it in a desired location.





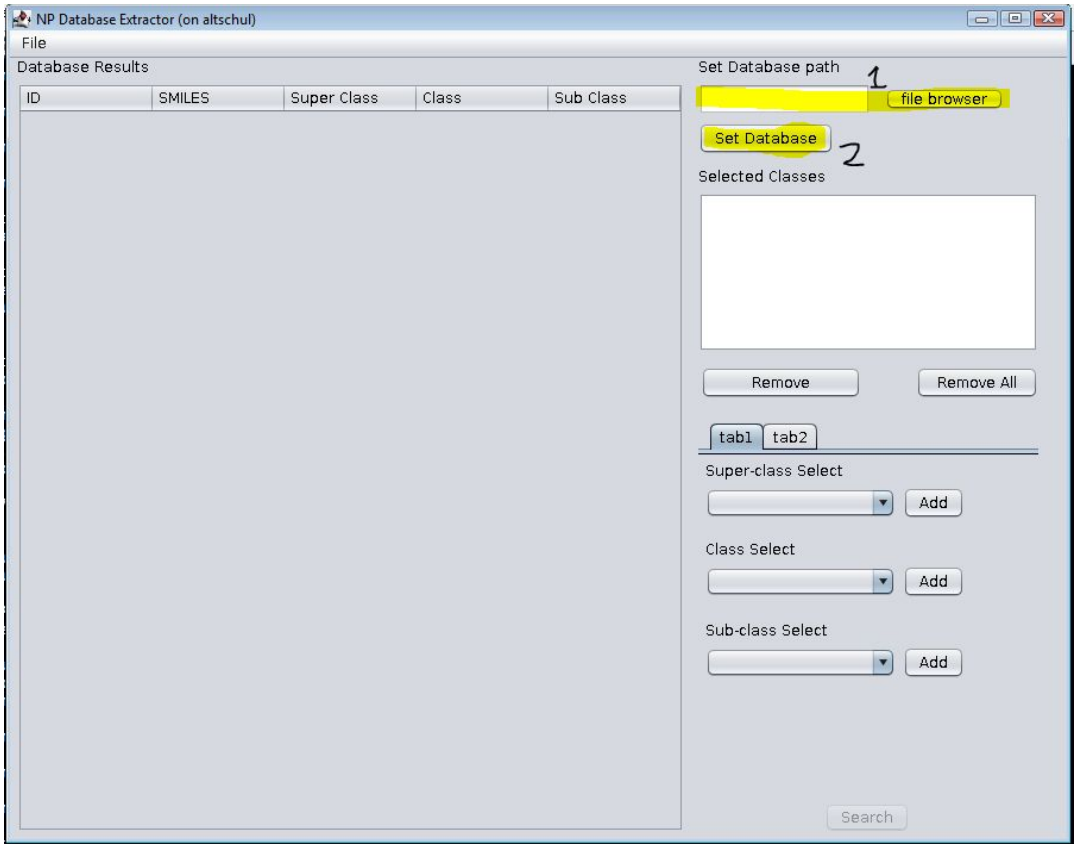
Editing the table name and columns

In case you have a different named table or the columns are named differently the values can be changed in the code at the top of the script NpDbGui.java where the tableName,classColumn,subClassColumn, superClassColumn, idColumn and smileColumn are defined. Change these values to your db’s columns and table names and the tool should work fine.



Running the program

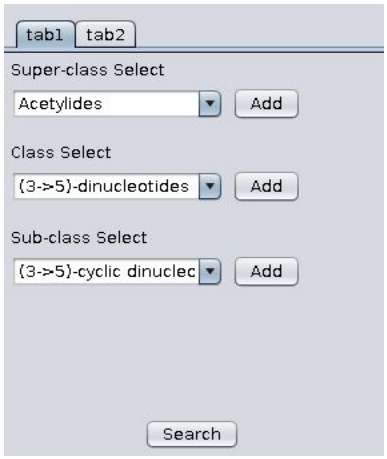
1. To run the program type: “java -jar /path/to/folder/NaturalProductsDbGui.jar” and press enter this should start the GUI.



2. Next select the database with the file browser or just enter the path in the textfield. Once done click the “Set Database” (2) button. This can take a minute to load.

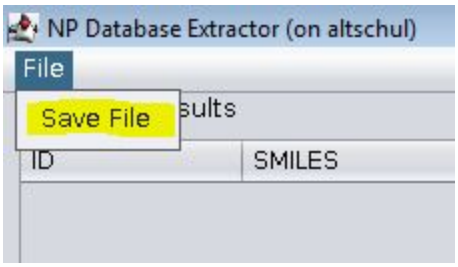
3. Once done the Search button should be clickable and both the tabs should contain dropdown menus filled with classes to select and a dropdown tree also containing the classes.

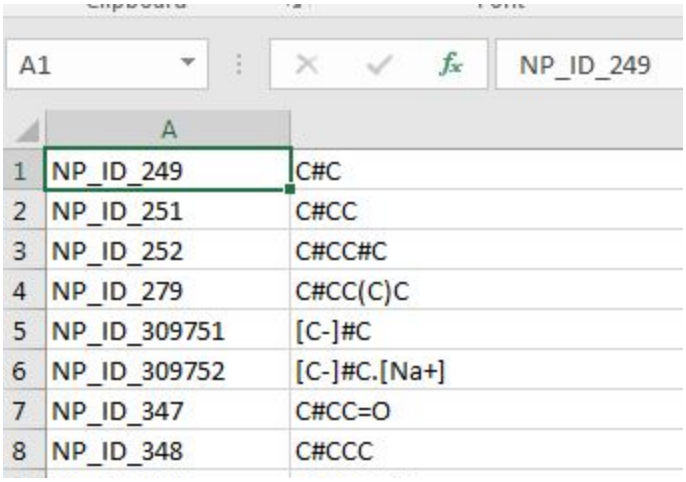
4.Select the classes you are interested in by clicking the add button once you have selected one. You can do this for as many as you want (keep in mind that the search time increases the more classes are selected).



5. Finally once done selecting class you can press the search button. This will start the search, once it’s done the results will show up in the table on the left. Here you can scroll through and review the results.

6. To Save the results click on file in the menu bar and click on the save button. Then select the path to which you want to save the file. The file will be saved with the entered name, make sure to either name it something ending in .txt or .csv. The output file should look something like the picture below, containing ID, SMILES pairs.





	A	
1	NP_ID_249	C#C
2	NP_ID_251	C#CC
3	NP_ID_252	C#CC#C
4	NP_ID_279	C#CC(C)C
5	NP_ID_309751	[C-]#C
6	NP_ID_309752	[C-]#C.[Na+]
7	NP_ID_347	C#CC=O
8	NP_ID_348	C#CCC