
Subject: Re: Filter out nasty functions

Posted by [nbehrnd](#) on Tue, 23 May 2023 05:09:10 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear Julio,

so far, I understand your approach as following: departing on e.g., selected-15B10Evolutionary_Library.879sentbefore.dwar, you load the 1.dwam macro to filter out compounds which are not good enough for further consideration. In the attempt to replicate this, at level of exporting the results as .sdf files (the macro still running), there are multiple error messages. This either could be a) because of your macro, b) because of the version of DW I use (DW for Linux including the updates packaged by 2023-05-18), or c) a combination of the two.

I briefly tinkered a macro attached below which does the the filtering, however requires manual intervention to save the results as .dwar and .sdf. Conceptually, it builds on DW's assignment of toxicity properties -- like the one you built. This however is followed by calculating a column with an if clause; if you would apply this manually: Data -> Add Calculated Values. As for the formula used: DW's "if syntax" basically is

```
if(test condition, positive case, negative case)
```

as the general form, and

```
if (Mutagenic == "none" && Tumorigenic == "none" && ReproductiveEffective == "none" && Irritant == "none" && NastyFunctions == "", "retain entry", "skip entry")
```

as the one used here. You recognize e.g. Mutagenic as one of DW's assigned functions; here, DW is requested to check if the assignment was negative (as expressed as the string "none"), yet simultaneously (the &&) was fine about Tumorigenic, ReproductiveEffective, and Irritant. For NastyFunctions, I opted for an empty string as condition for a compound useful to retain. This is why, if an entry passes all these five tests well, its corresponding entry in the new column to build will be the string "retain entry", or else "skip entry".

The script then removes the filters of the individual properties (Tumorigenic, Irritant, etc) to leave only one if the compounds are good, or not for further work; for the ease of work with an additional green, or red background of the cell. For a much smaller set of molecules, this macro works well enough and does not yield the errors I observed with your macro. It however does not (yet) automatically save the results of «filtering» in a separate .sdf/.dwar file. Is this approach in line of what you like to accomplish?

As for «what defines a function nasty», one would have to check the source code,[1] as e.g. file /src/com/actelion/research/datawarrior/task/chem/DETaskCalculateChemicalProperties.java for example contains the string «nasty» 16 times and how this property is assigned either in this file's functions, or elsewhere in the source code. Maybe some criteria to mark compounds as not well suitable are similar to the ones in Lilly's criteria.[2]

Norwid

[1] <https://github.com/thsa/datawarrior>

[2] <https://github.com/lanAWatson/Lilly-Medchem-Rules>,

https://github.com/lanAWatson/LillyMol_6_cmake

File Attachments

- 1) [20Random_Molecules.dwar](#), downloaded 264 times
 - 2) [suitable_compounds_color.dwam](#), downloaded 269 times
 - 3) [20Random_Molecules_processed.dwar](#), downloaded 257 times
 - 4) [2023-05-23T06.26.35 -- screenshots.png](#), downloaded 190 times
-