Subject: Comparing Structures from two Datawarrior Files
Posted by ELFritzen on Thu, 29 Jun 2023 14:35:47 GMT
View Forum Message <> Reply to Message

I have two Datawarrior Files that I would like to compare the structures. One file is our corporate database, containing all the compounds we have prepared and registered. This file contains 729 compounds. The other file contains structures that have been enumerated using Datawarrior and contains 690 compounds. The enumerated file contains structures that have already been made and registered and I would like to find these structures and eliminate them from the enumerated file, so as not to target them for synthesis again. I've tried using the "Find Similar Compounds in File...." command, but Datawarrior locks up with the display of many "Uncaught Exception:Null" messages being produced. I end up having to force quit Datawarrior. I've tried using either the "Structure of SMILES{FragFp] or the Structure{Exact] options for the Structure Column[comaparison method]: and get the same result. The comparison starts, and eventually freezes Datawarrior. Is there something I'm missing? Or is there a better way to do this?

Thanks in advance for help.
Edward Fritzen

Subject: Re: Comparing Structures from two Datawarrior Files
Posted by zig on Mon, 03 Jul 2023 18:29:40 GMT
View Forum Message <> Reply to Message

Dear Edward,

a work around could be to create a unique code for your molecules, with Chemistry>From Chemical Structure > Add Canonical Code..., in both files.

Then you select all these canonical codes in the file containing the molecules of the corporate database, and copy them (Edit> Copy -> they are now in your clipboard)

Switch to the file with enumerated molecules and create a list with the canonical codes copied in your clipboard (List>Create Row List From > Clipboard)

A List Membership filter is created on the right side. You can now hide the molecules that were in the first file.

I hope this helps. :) Another way would be to merge both files using the canonical code as merge key.

Subject: Re: Comparing Structures from two Datawarrior Files
Posted by ELFritzen on Thu, 06 Jul 2023 12:51:10 GMT
View Forum Message <> Reply to Message

Hi zig,

Thanks for your reply and helpful tips. I've found a way to do what I want by merging the two files using structure as the merge key, and by selecting the option to append rows for the compounds not contained in one of the files. I think that works as well.

Ed

Subject: Re: Comparing Structures from two Datawarrior Files
Posted by thomas on Sun, 16 Jul 2023 09:31:30 GMT
View Forum Message <> Reply to Message

Hi Ed,

I tried the comparison with a couple of files with version 5.5.0 and with the current dev build and could not reproduce the "Uncaught Exception:Null" error. If you get the error with the current dev version. then I would be glad, if you could send me files and exact procedure to reproduce, which I need to detect and remove the bug.

Many thanks in advance, Thomas

Subject: Re: Comparing Structures from two Datawarrior Files
Posted by ELFritzen on Mon, 17 Jul 2023 15:06:37 GMT
View Forum Message <> Reply to Message

Hi Thomas,

Thank you for your help. I'm a little hesitant in sharing the files, since they contain proprietary compounds than have not been covered by a patent. I was able to devise a work around as described above. I'll think about it a little more to see if I can provide some files that produce the error I see.

Ed
