Subject: Comparing two databases by merging
Posted by tingjenc on Mon, 17 Jun 2024 03:06:52 GMT
View Forum Message <> Reply to Message

Hi,

I tried to compare two SDF databases by merging two databases with structure as the merge key. I also followed a previous discussion by using canonical code for merging. All I got is "The defined key column(s) contain duplicate data in some rows and cannot uniquely identify each row." I indeed used ChemFinder to confirm that there are no duplicated structures in the two databases. So, I am stuck here. Is there anything I can do to use "merge" successfully? Or I can try other ways to compare two databases?

Subject: Re: Comparing two databases by merging
Posted by thomas on Fri, 21 Jun 2024 08:05:50 GMT
View Forum Message <> Reply to Message

Hi tingjenc,

Structures are stored as canonical text string (idcodes). Thus, unless you intend to merge different stereo isomers or tautomers, you don't need to use canonical codes for merging. When merging, only the second file's key column(s) need to be unique. You could try a couple of things:

- change the order of your files. Possibly, one of the two files has unique structures.

- before merging the second file, you could "Data -> Merge Equivalent Rows" selecting 'Structure' as criterion. After that merging both files using 'Structure' should work.

- To find and display your duplicate structures within the second file you could "List -> Create Row List From -> Unique Rows" using the 'Structure' column. Then, select 'unique rows' in the new list filter, invert the filter, and click on the 'Structure' table header to sort by Structure. Redundant structures are now shown together.

- Instead of merging, you could use 'Chemistry -> Find Similar Compounds In File...". Here you can define a similarity limit using any descriptor rather than using an exact structure match, what merging does.

Hope, this helps...

Subject: Re: Comparing two databases by merging
Posted by tingjenc on Thu, 11 Jul 2024 08:26:45 GMT
View Forum Message <> Reply to Message

Thank you very much. The suggestions work great.