Subject: Re: Assign cluster name based on cluster size
Posted by nbehrnd on Fri, 22 Apr 2022 20:46:45 GMT
View Forum Message <> Reply to Message

Hello mcmc,

I just completed a small Python script to process DataWarrior's results about structure similarity (Chemistry -> Cluster Compounds) exported as text file (File -> Save Special -> Textfile). It identifies the clusters, sorts these based on the number of molecules in each clusters, updates the molecules' cluster labels (1, 2, 3,...) accordingly and writes a new .txt file one may read with DW by (Ctrl + O). There are two sorts possible: a) «the more molecules in the cluster, the lesser the integer used as label of the cluster», a pattern possibly matching best your intent. Though with the optional flag -r you equally may reverse the sort for b) «the more molecules in the cluster, the greater the label».

The .zip archive attached below includes the .py script and describes early results when processing a small set of test data. It assumes the first column labeled «Cluster No» contains the cluster labels assigned by DataWarrior (which is the program's default header).

Norwid

File Attachments

1) 2022-04-26_datawarrior_clustersort.zip, downloaded 247 times