Subject: Re: Count # scaffolds by plate ID
Posted by nbehrnd on Thu, 06 May 2021 12:24:18 GMT
View Forum Message <> Reply to Message

It is possible to use the structures of one .dwar file for a comparison against structures in a second .dwar file. Because I did not identify (yet) how to point from the frequency table generated specifically to the SMILES strings of the Murcko scaffolds of the original file, I wrote a little Python script to address the initial question. The script depends on standard CPython and the RDKit library only. For a small library (say, .le. 3k entries), I perceive the rate of computation as fast enough; of course your mileage may vary when submitting 500k.

The archive, in addition to the script, equally documents a test run with said script.

Norwid

File Attachments

1) scaffold_seeker.py, downloaded 2277 times
2) scaffold_seeker.zip, downloaded 754 times