Subject: Re: Why is clustering limited up to 10000 molecules?
Posted by thomas on Thu, 23 Jul 2015 19:40:59 GMT

Clustering is a very old functionality and not used very often. The used algorithm is reproducible and is not based on random starting point as other faster algorithms. The flipside, however, is that is not a particularly fast one and its memory consumption and processor usage grow quadratic with the size of the data file. Without a limit people would experience error messages or extreme processing times causing lots of frustration. I am aware that DataWarrior needs an alternative clustering algorithm and it is on the feature list for the future, but till now I considered clustering low priority, because often there are better methods to reach a certain goal.

Kind regards, Thomas