Subject: How are compounds clustered?
Posted by mindhira on Fri, 13 May 2016 15:09:13 GMT
View Forum Message <> Reply to Message

I'm using Datawarrior to cluster various sets of compounds but it would be helpful if I could tell people how these are clustered. From the description, it sounds like hierarchal clustering. Is that accurate?

Subject: Re: How are compounds clustered?
Posted by thomas on Fri, 13 May 2016 19:22:10 GMT
View Forum Message <> Reply to Message

You are right. It is an agglomerative hierarchical clustering. It starts by calculating the complete similarity matrix between all molecules. Therefore it can be used by any descriptor, not just vector based ones. At the beginning every molecule already represents a cluster. Stepwise always those two clusters are merged, which are the most similar ones. The respective two rows and columns of the similarity matrix are then also merged by simply calculating a weighted mean between any two similarity values. The merging continues until the stop criterion is met.

Thomas