Subject: Clustering by similarity of srtucture
Posted by Rasajna on Mon, 21 Apr 2025 18:55:23 GMT
View Forum Message <> Reply to Message

Hello, I was using data warrior for clustering molecules from a database based on the structure. I wanted to know more about the parameters and the algorithm being used so I can explain it while presenting my work and just better understand the results. I saw an answer on the forum that said DW uses agglomerative hierarchical clustering but what is the the measure of similarity, is it tanimoto's coeffiecient?

Subject: Re: Clustering by similarity of srtucture
Posted by thomas on Mon, 21 Apr 2025 19:11:41 GMT
View Forum Message <> Reply to Message

that depends on the descriptor you use. FOr the simple binary fingerprints (FragFp, PathFp, and SphereFp) it is Tanimoto. SkeletonSpheres similarioty is also a kind of Tanimoto, but it is based on fragment count numbers rather than fragment existence bits. OrgFunctions and Flexophore similarities are the result of graph matching algorithms, the first based on best matching trees and the second on an MCS matching of complete graphs. The Flexophore is published and some basic information about the descriptors can be found here: https://openmolecules.org/help/similarity.html.