Subject: Re: Developing plugins using SDK2
Posted by thomas on Wed, 13 Jul 2022 20:19:44 GMT
View Forum Message <> Reply to Message

It seems you refer to example 2. Please provide some more light on your environment. I do not fully understand the question. It seems that you have some kind of database that contains molfiles. Does the database support structure searches (similarity and substructure)? If yes, then your code should make sure that only structure search hits are retrieved. If there is no way to search on the database, you need to retrieve all molfiles, which naturally will take a long time. In this case retrieval is the bottleneck. You could filter in the plugin code to only pass structure search matches to DataWarrior. But much easier would be to just add all structures to a DataWarrior table and to use a structure filter to do the filtering.

The best option, however, would be to ask your server to do the structure search and to only retrieve structure hits over the network. Second, molfiles are very large. Retrieving a compressed list of molfiles or a compressed SD-file would be an easy solution. Better would be to use a more compact stringified format. If you can live with the stereo deficencies, you could use SMILES. Better would be to add OpenChemLib idcodes into the database and to retrieve them. They are much more compact than SMILES, are canonical and contain the full enhanced stereo representation.