Subject: Developing plugins using SDK2
Posted by helloWorld22 on Wed, 13 Jul 2022 07:17:59 GMT
View Forum Message <> Reply to Message

I'm a new user to datawarrior and trying to implement a plugin supporting structure search on our internal DB.

I have already downloaded SDK 2.0 for dw and stuck at run() where queryMolFile and similarityLimit is captured from the configuration - CONFIGURATION_QUERY_STRUCTURE and CONFIGURATION_SIMILARITY_LIMIT.
After these values are captured, how can I search a DB file containing mol files at the run time? The mol files are stored into DB and I'm retrieving them using a query and storing them in rs and then showing them in the pluginHelper object in the cell with column type - 2.
So, I can view all the 1000s of mol files but I cannot filter them at run time based on similarity and user query structure.

Can you please help me how I can search with the queryMolfile and similarityLimit?

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.

Subject: Re: Developing plugins using SDK2
Posted by helloWorld22 on Thu, 14 Jul 2022 14:04:54 GMT
View Forum Message <> Reply to Message

Hi Thomas,
Thanks for replying.

Yes, I was referring to example2.

Currently, the DB contains mol file and SMILES as CLOB and string in the DB, respectively.

I want to do a similiarity search but dont want to get everything from the DB and then filter it.
So what I want to do is:
Use Datawarrior or openchemlib.jar to do a similarity search on existing database.

I believe, since datawarrior is a client, it needs all the data from DB to do the similarity search.

Is it possible to achieve the same feature of datawarrior bundled with openchemlib.jar and implement it in an API?
