Subject: Can you disable SMILES conversion?
Posted by zsolt on Mon, 07 Dec 2020 11:40:06 GMT
View Forum Message <> Reply to Message

a quick question on SMILES conversion: using DataWarrior with some larger datasets we noticed that SMILES conversion takes an unacceptably long time. For certain workflows it is not required for us so would like to ask it it is possible to disable it?

Another idea I am thinking as a workaround is to somehow craft a DataWarrior specific file (e.g. a .dwar) where I specify the columns containing my SMILES as plain text type - maybe then DataWarrior won't try converting it. (However I would not like to have to specify all the extra bits that are contained in the .dwar file like the position of the windows, filters etc...)

Thanks in advance,
Zsolt

Subject: Re: Can you disable SMILES conversion?
Posted by thomas on Tue, 08 Dec 2020 16:30:29 GMT
View Forum Message <> Reply to Message

If you just give your TAB delimited text file the '.dwar' extention, then DataWarrior doesn't try to convert SMILES codes. A content sample would be: (your file could even contain chemical structure as idcodes, which would be recognized, if the column name ends with ' [idcode]'

ID Smiles
1 CCOCC
2 CCCO
3 CCO

Subject: Re: Can you disable SMILES conversion?
Posted by zsolt on Mon, 14 Dec 2020 23:32:07 GMT
View Forum Message <> Reply to Message

Thank you very much for the reply - this works, the processing is much faster indeed.

The only issue now is that when I try to set these plain SMILES strings to be on one of the axes of the 2D plot I get the error message "Column <column name> is neither a descriptor nor does it have a numerical variance" if I try to use it from a macro and it is missing from the options if I try to set it through the GUI. It makes sense not to be able to set a string as one of the axes however it seems to be possible for other string/plain text type columns - presumably because they are "descriptor"s. Is there a way for these plain-text-SMILES columns to be specified as an axis of the 2D and 3D plots?

Thanks,

Zsolt

Subject: Re: Can you disable SMILES conversion?
Posted by thomas on Sat, 19 Dec 2020 11:36:08 GMT
View Forum Message <> Reply to Message

In this case the 'Smiles' column is a normal text column, for which DataWarrior keeps a sorted list of available categories unless the number of different text entries exceeds 65536. As long as the number of distinct smiles values (categories) stays below that value, the column is considered to contain categories and can be assigned to an axis. Categories have a natural or user-selected order and, therefore, individual smiles can be assigned a numerical value (its position in the category list), which is used to position it on the axis. If the number of distinct values in a column is larger than 65536, then DataWarrior doesn't maintain a category list anymore. Then it cannot assign numerical values to individual smiles to position them on an axis. For efficiency reasons there must be a limit.