Subject: SMILES Code
Posted by Anne on Tue, 12 Sep 2023 07:27:49 GMT
View Forum Message <> Reply to Message

When I paste the SMILES codes of some structures, expecially the one with stereochemistry indication, an error message "Uncaught Exception:2" would pop out more than one hundred times. I can only press "OK" more than one hundred times to bypass this error. :?   :cry:

D-Glucal is an example. I pasted the SMILES code "O[C@H]1[C@H](O)[C@@H](CO)OC=C1" into the structure drawing panel, and then the error message shown again and again.

My DataWarrior version is V05.05.00. Could you let me know how to fix this error? Do I miss any step?

Subject: Re: SMILES Code
Posted by thomas on Wed, 13 Sep 2023 15:53:05 GMT
View Forum Message <> Reply to Message

I tried to reproduce the problem, but failed. When I paste the SMILES from your message I always got the proper molecule or substructure, depending on the context. I tried directly in the table view, in a structure filter (directly and after opening the editor). I also tried in the Chembl retrieval dialog in substructure and similarity modes. I also tried with the old V5.5.0 as well as with the current development version.

There must be a different reason. What is the source of the SMILES? May there be invisible characters in between? Can you help me to reproduce the error? Which OS do you use? I suggest to update to the dev version (go to DataWarrior download page, click on 'read and understood..., and find the update link in the small print).

Subject: Re: SMILES Code
Posted by Anne on Mon, 18 Sep 2023 08:55:50 GMT
View Forum Message <> Reply to Message

Thank you for your kind and prompt reply. The SMILES code is transformed via ChemDraw. I can also obtain the structure in DataWarrior via that SMILES code, but it seems the error message is unavoidable. I need to press "OK" to make the system accept the structure. However, there is no error message when I paste the SMILES code "C1CCCCC1C2CCCCC2", which corresponding to a structure without any stereochemistry information.

We install the DataWarrior on Win10, traditional Chinese version. Would this error caused by the conversion between double-byte and single-byte system?

Subject: Re: SMILES Code

Posted by thomas on Tue, 19 Sep 2023 21:01:29 GMT
View Forum Message <> Reply to Message

I assume that you a right and the problem is connected to the Chinese Windows version, on which the default character encoding is probably GBK, while on western OS version it is UTF-8.

I am currently checking a few things simulating with standard encoding GBK, but don't understand the cause of the problem yet...

Subject: Re: SMILES Code
Posted by thomas on Wed, 20 Sep 2023 11:03:36 GMT
View Forum Message <> Reply to Message

I assume that the problem happens, when the String from the clipboard is converted to bytes before the SMILES is parsed. In version 5.5.0 the conversion of the String to bytes is done by the default operating system character set. I simulated that using various Chinese default character sets as GBK, Big5, and GB18030. It seems that all these convert the SMILES nicely to one byte per character, because all the characters in the SMILES (including '@') seem to be represented in all these char-sets by the same one-byte value as when using UTF-8. However, when using UTF-16 every character is converted into two bytes, which causes as lot of trouble. Thus, I assume that your machine default character encoding is an unusual one that does not convert '@' into one byte.

I have update DataWarrior to explicitly use UTF-8 for most of the String<->byte[] conversions, which should also cover the clipboard stuff. Can you please check with the current dev version (dw_wl_win.zip), whether the problem is solved? If not, I suggest to download and use the dw_wl_win_d.zip update file (same URL plus '_d'). If you start the contained DataWarrior_d.exe from the command line, you get debug output that hopefully explains, if sent to me, where to look for the problem.

Sorry for the trouble,

Thomas

Subject: Re: SMILES Code
Posted by Anne on Thu, 21 Sep 2023 02:06:13 GMT
View Forum Message <> Reply to Message

:o :o :o Thank you so so much for your reply. I have sent the examples to you. Sorry that I spend some time to make this clear.

While I use "paste special" to transform the data from EXCEL, the SMILES code transform to each corresponding structure successfully without message.

While I try to add a whole new structure to the DataWarrior after the paste special step, countless messages will pop out. I need to correct my previous report. Regardless of the stereochemistry,

the unexpected message will show. "CC(=O)C" also has the same outcome.

When I paste the "O[C@H]1[C@H](O)[C@@H](CO)OC=C1" to different conditions, different situations will show.

Please feel free to let me know any idea from you. I can try.

Subject: Re: SMILES Code
Posted by thomas on Fri, 22 Sep 2023 09:43:19 GMT
View Forum Message <> Reply to Message

Many thanks for the examples. I could finally reproduce the problem. It wasn't related to character sets. The culprit was the 3D-view, which, when showing molecular structures on an axis, forgot to increase the molecule buffer space, when new molecules are added through pasting them directly into the table.

I have fixed this and deployed the update as dw_wl...zip archives with replacement files, which can be downloaded as decribed above. Please let me know, if something is not working as expected...

Subject: Re: SMILES Code
Posted by Anne on Mon, 25 Sep 2023 05:22:12 GMT
View Forum Message <> Reply to Message

:lol: It works! 8o
I sincerely appreciate all your time and effort for this. :o   Please feel free to let me know if there is any further step I should take.