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