Subject: > and < signs not interpreted in 6.0.0
Posted by mcmc on Thu, 11 Jan 2024 11:19:55 GMT
View Forum Message <> Reply to Message

I noticed that 6.0 treats values like <2 and >10 differently from 5.5
In the old version these were treated as numbers, and a column would be correctly assigned as numeric if it would contain entries like these. Allowing for rounding, colouring scales etc.
Now columns with >2 are interpreted as text, generating errors on all my old macros.
Manually setting the columns to floating point generates errors too (image attached).

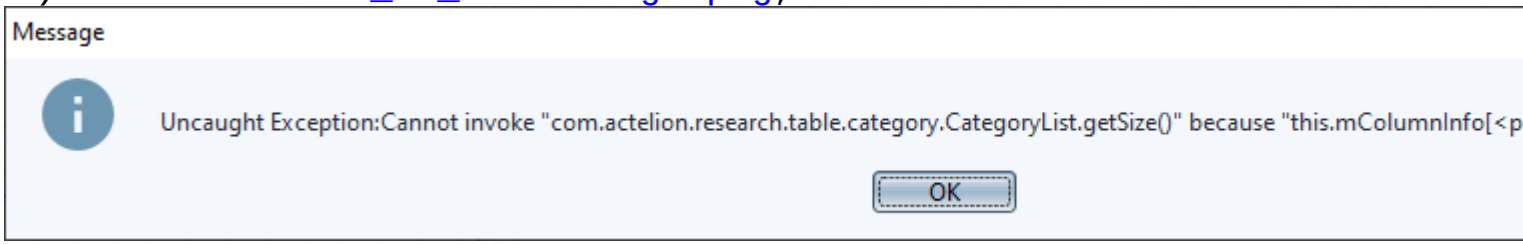
EDIT: there seems to be an overarching phenomenon, which is the treatment of spaces (notably the lack thereof) in cells.
'>2' seems to be recognized as text, while
'> 2' is treated as a number.
This was different in 5.5

Likewise, multiple numbers without spaces, are treated as text:
'4.5;4.8' is text, while
'4.5; 4.8' are two consequetive numbers.

Clearly it would be great it this could be backwards compatible to 5.5 again.

File Attachments

1) 2024-01-11 12_16_26-Message.png, downloaded 896 times



Subject: Re: > and < signs not interpreted in 6.0.0
Posted by nbehrnd on Thu, 11 Jan 2024 22:42:56 GMT
View Forum Message <> Reply to Message

Hello mcmc,

in a small library of random molecules generated by DW 6.0, I observe a close `<2`(example metal atom count), or `>30` (example nonH atom count) to be recognized as a comparison against a number; just as anticipated and just as the pattern experience prior to the transient to the new version.

Regards,

Norwid

File Attachments
1) Random_Molecules.dwar, downloaded 531 times
2) nonH_count.dwam, downloaded 605 times
3) metal_count.dwam, downloaded 1980 times
4) count_with_space.dwam, downloaded 1951 times
5) Random_Molecules_applied.dwar, downloaded 641 times

Subject: Re: > and < signs not interpreted in 6.0.0
Posted by mcmc on Fri, 12 Jan 2024 11:03:20 GMT
View Forum Message <> Reply to Message

Unfortunately, I cannot test this anymore, as I have rolled back to 5.5 now.
I had a DW file, with only a few rows, where one column consisted of these values
>2000
55.6
20
Right-clicking on the column heading indicated that these were not numbers. No option to round, for example.
As soon as I introduced the space by editing the cell to '> 2000', the column reverted to numbers and I could round.
This dwar file was not created from scratch with 6.0. Not sure if that could have been part of the problem, but one would still expect backwards compatibility to open older files.

The issue with the multiple numbers was reported by a colleague of mine, so we are at least two people observing space issues.

Subject: Re: > and < signs not interpreted in 6.0.0
Posted by nbehrnd on Sat, 13 Jan 2024 19:18:59 GMT
View Forum Message <> Reply to Message

In continuation with DW 06.00.00, I created a new array: the first column with three random molecules (DW generated), the second column added with Data -> Add empty columns, where I selected "text" as category. Manually, I added the strings `>2000`, `55.6`, and `20` to the cells as a test property (cf. attached .dwar and `dw_screenphoto_01.png`). The subsequent addition of a third column by Data -> Add Calculated Values used a function of

if(test_property>20, "big", "small")

proceed smoothly. In your case application, do you equally have the automatic assignment of the column type enabled? This is based on the observation the intended computation is going to fail if the second column (still/accidentally) is set to the level of `text` -- regardless if `force categories` is activated, or not. Subsequent adjustment of the second column, then re-run by right-click on the

third column and "update formula and recalculate" adjusted the results.

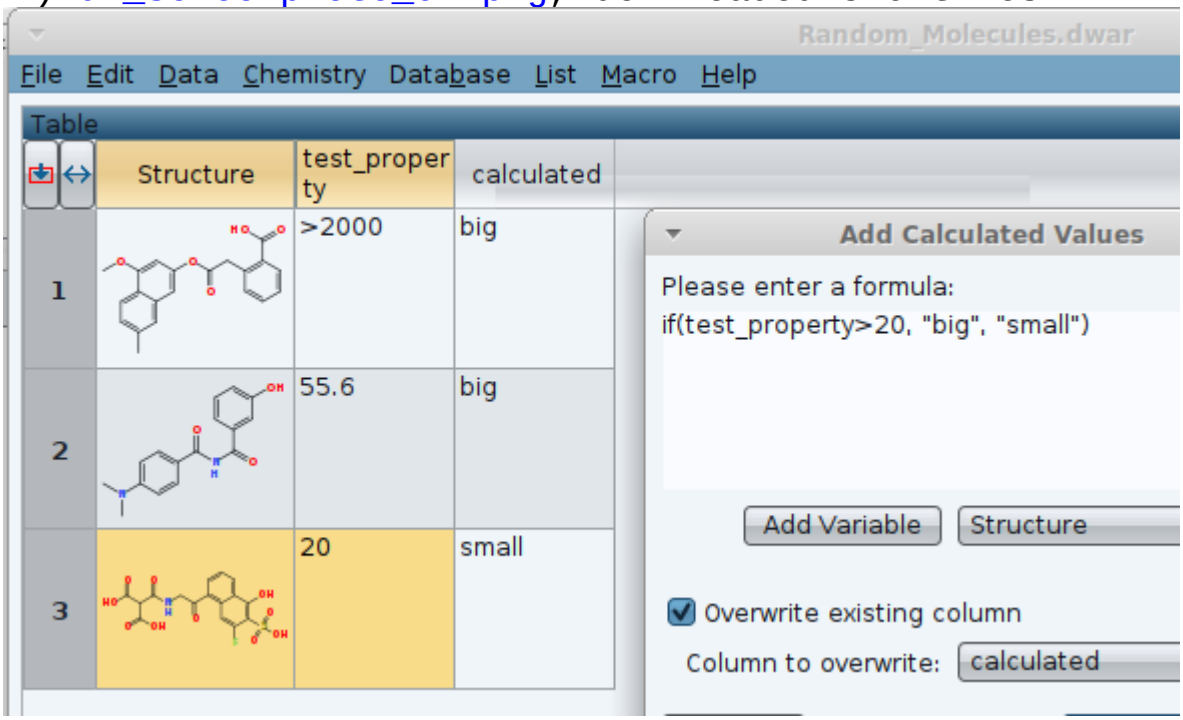
However: I assumed a right-click on the third column and subsequent "Re-Calculate All Columns" would be more helpful for a global update of the array of maybe multiple columns of calculated values. Hélàs no, it replaced the original entries in the second column by the results of my formula, and yielded the "NaN" indicator in the third column (dw_screenphoto_03.png). This observation is less intuitive to me.

Regards,

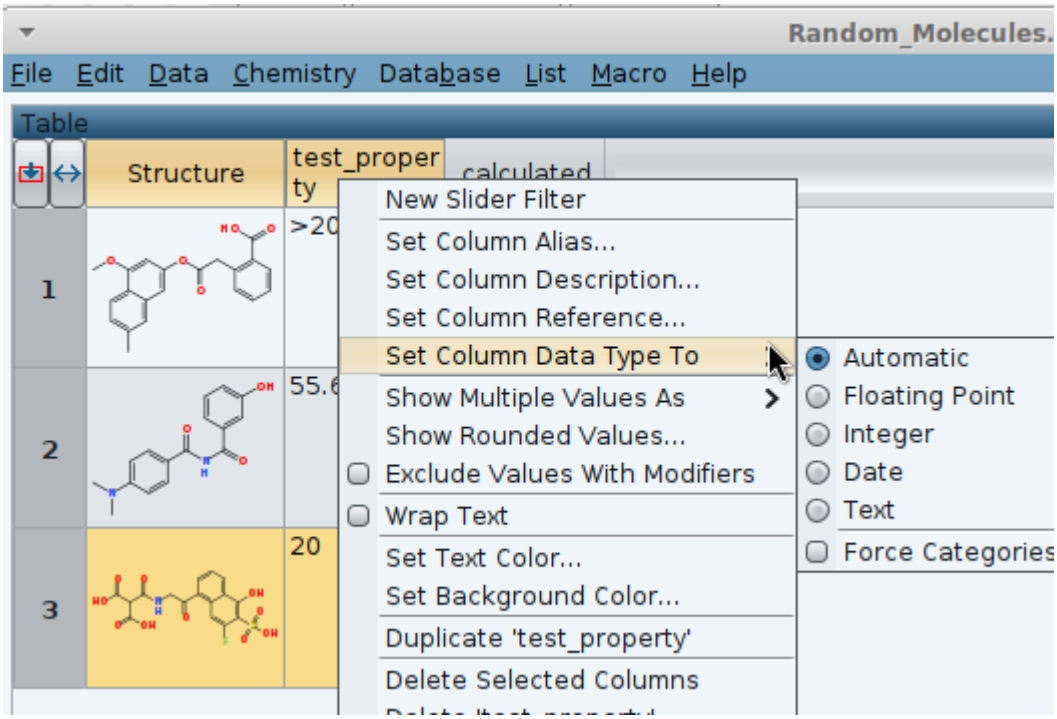
Norwid

File Attachments

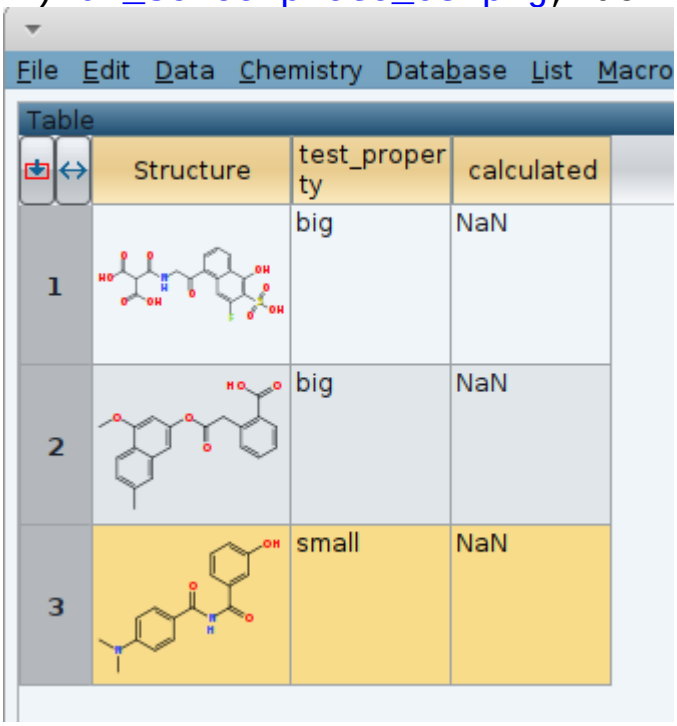
1) Random_Molecules.dwar, downloaded 596 times
2) dw_screenphoto_01.png, downloaded 840 times



3) dw_screenphoto_02.png, downloaded 761 times



4) dw_screenphoto_03.png, downloaded 836 times



Subject: Re: > and < signs not interpreted in 6.0.0
Posted by mcmc on Wed, 17 Jan 2024 09:18:57 GMT
View Forum Message <> Reply to Message

Hi Norwid, all columns are set to automatic type. The same .dwar file behaves differently in 5.5

and 6.0 in terms of interpretation of these spaces. Numbers become text - presumably due to the automatic type.

Subject: Re: > and < signs not interpreted in 6.0.0
Posted by thomas on Sat, 20 Jan 2024 14:02:51 GMT
View Forum Message <> Reply to Message

To mcmc: DataWarrior 6.0.0 should be compatible with earlier versions regarding the interpretation of numbers with modifiers like '>2000'. I tested your example (>2000, 55.6, 20) which correctly recognized the column as numerical. Thus, I assume there is something else fishy. Which OS do you use? Possibly the OS localization setting has some influence. Do you use decimal points '.' or ','? If I can reprodice, I may be able to update the numerical recognition circumvent the problem.

To Norvid: Many thanks for your comments on this. I could reproduce your phenomenon with writing into the 'test_property' column instead of the 'Calculated' column. This happened when clicking 'overwrite' twice. When changing the overwrite setting, DataWarrior was adapting the target column setting in a strange way: when choosing to overwrite, it hanged the setting to an existing column ('test_property') and when unchecking again, it was not updating the target column. It was also not perceiving that 'not to overwrite' an existing target column is not really compatible. I updated the behaviour to be more meaningful.

Subject: > sign properly interpreted, but '2;3' is interpreted as text, not numercal
Posted by mcmc on Tue, 30 Jul 2024 10:11:36 GMT
View Forum Message <> Reply to Message

Hello. Finally getting back to this issue now.
I am running 6.2.1 on my private PC now, and can reproduce half of the issue encountered earlier.
That is, the < and > signs are correctly interpreted, also without a whitespace in between.

However, two numbers in one cell, separated by nothing but a ; is not interpreted as numerical.

Started a fresh DW session and loaded the Wikipedia compounds.
Hovering over the Molweight column header, it nicely shows the values in there are numerical.
This is still correct if I add '>' to the first value (image1).
However, if I add a second number to a molecular weight, without whitespace, the perceived data type changes to text (image2). Adding a space reverts it back to numerical (image 3).

I also attach the .dwar (only first 3 wiki compounds).

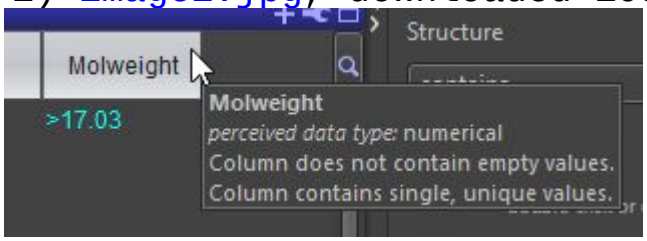
(the reason this is important to us, is that we use an in-house application that generates dwar files from our internal project SQL databases. If say an IC50 was measured more than once, each

value is written into the dwar file separated by ';' without whitespace. We can of course change those scripts to include the whitespace, but we have hundreds of dwar files in circulation without the whitespace, that cannot be properly interpreted anymore by DW6).

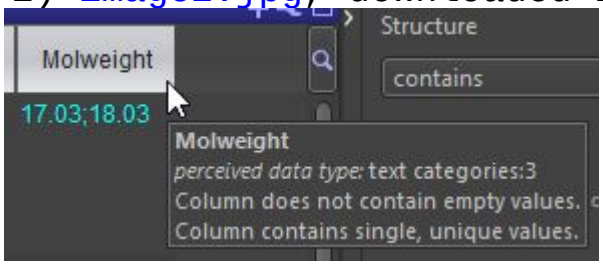
[EDIT: 6.2.1 tested on Win11 Home 23H2. Initial report was with 6.0.0 on Win10 Pro, likely 22H2]

File Attachments

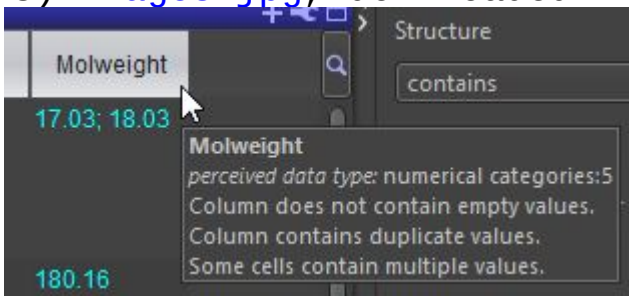
1) image1.jpg, downloaded 1096 times



2) image2.jpg, downloaded 1167 times



3) image3.jpg, downloaded 1137 times



4) Wikipedia_Compounds.dwar, downloaded 466 times

Subject: Re: > sign properly interpreted, but '2;3' is interpreted as text, not numercal
Posted by nbehrnd on Tue, 30 Jul 2024 17:31:32 GMT
View Forum Message <> Reply to Message

Dear mcmc,

with multiple measurements of one property about one structure record, could 1) creating individual .dwar files (each about one structure record with exactly one measurement of the property of interest) and 2) the subsequent merge of equivalent rows (where multiple measurements in one cell cell are separated by a linefeed) represent a workflow suitable for you? Though I didn't yet use DW in the context of SQL, your question seems to relate to the Yuan's

approach documented by May 28th, 2024.[1]

Best regards,

Norwid

[1] https://openmolecules.org/forum/index.php?t=msg&th=743&a mp;start=0&

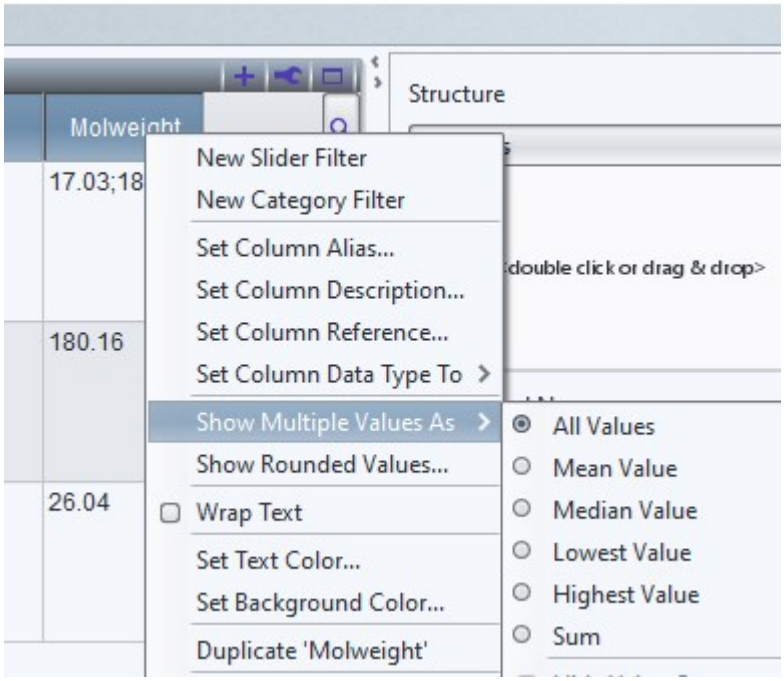
Subject: Re: > sign properly interpreted, but '2;3' is interpreted as text, not numercal
Posted by mcmc on Wed, 31 Jul 2024 07:52:50 GMT
View Forum Message <> Reply to Message

Thanks for the reply, Norwid. I take it you could reproduce the issue? Also on windows?
DW has clearly been designed to handle multiple values in one cell (see image).
In plotting, it will automatically plot the average, which is typically what you want.
In DW5, there was no difference in handling '2;3' vs '2; 3'.

There is some resemblance to Yuan's request (although I would never put results from different assays into the same cell), but that doesn't solve my backwards compatibility issue..

File Attachments

1) image4.jpg, downloaded 1088 times



Subject: Re: > sign properly interpreted, but '2;3' is interpreted as text, not numercal
Posted by thomas on Wed, 31 Jul 2024 12:51:36 GMT
View Forum Message <> Reply to Message

Dear mcmc,

sorry, it took me a while to understand the issue. A single semicolon was never meant to be a valid separator. DataWarrior always creates and expects '; ' (and '\n'). Nevertheless, you are right that version 5 also separated entries, when delimited with a single semicolon. This was more an error than intended.

For compatibility I will go back to the old handling in version 6.2.2 that will be out soon...

Sorry for the inconvenience,

Thomas

Subject: Re: > sign properly interpreted, but '2;3' is interpreted as text, not numercal
Posted by mcmc on Wed, 31 Jul 2024 15:50:28 GMT
View Forum Message <> Reply to Message

Excellent - thanks a lot!
