Subject: Splitting column data
Posted by sansun on Mon, 27 Jul 2020 12:20:13 GMT
View Forum Message <> Reply to Message

I have a column with references in the following format.

Proc. Natl. Acad. Sci. U.S.A. 105 (26), 9059-9064 (2008)

Is it possible to extract 'journal name', 'year', etc. in separate columns?

Subject: Re: Splitting column data
Posted by thomas on Wed, 29 Jul 2020 08:50:49 GMT
View Forum Message <> Reply to Message

if your references would be in a column called 'Lit' you could use something like the attached macro. The problem is that there is no easily recognizable separator between the journal name and the first number. This could be solved with a new function lastIndexOf() or reverseIndexOf(). Please let me know, if your references are similar enough for the macro to work. Then I would add the needed function for the remaining cut to make.

File Attachments
1) SplitLit.dwam, downloaded 653 times

Subject: Re: Splitting column data
Posted by sansun on Wed, 29 Jul 2020 10:19:35 GMT
View Forum Message <> Reply to Message

Hi Thomas,

Yes references are in a similar format. However, in some cases several references are there in the same cell due to merging of rows.

Mainly I need 'year' information which is within the brackets.

Subject: Re: Splitting column data
Posted by sansun on Wed, 29 Jul 2020 10:26:31 GMT
View Forum Message <> Reply to Message

I tried your macro and following is the result.
I have several references in a cell and it recognizes only the first one, I guess.

Subject: Re: Splitting column data
Posted by sansun on Thu, 30 Jul 2020 05:28:59 GMT
View Forum Message <> Reply to Message

Another observation related to this query.

When I download Chembl data from within Datawarrior (from Database tab), it gives references in the above-mentioned format in a single column.

However, if I directly download data from the Chembl website in the form of .sdf file it gives 'journal name', 'year' etc. in the separate columns.

Probably the later format is better if splitting is going to be a problem.