Subject: Retrieve Wikipedia Molecules
Posted by MedChemProf on Thu, 02 Apr 2020 16:25:23 GMT
View Forum Message <> Reply to Message

Dear DataWarrior Forum Members,

I would like to start using the "Retrieve Wikipedia Molecules" function in the Database options of DataWarrior. I am certainly not a database wizard and don't have any experience writing search scripts for databases. Is there either a way to tell or a description of the "Retrieve Wikipedia Molecules" option? I am trying to understand specifically what it is looking for when it queries Wikipedia.

Thank you, Chase

Subject: Re: Retrieve Wikipedia Molecules
Posted by nbehrnd on Fri, 03 Apr 2020 13:51:00 GMT
View Forum Message <> Reply to Message

Dear Chase,

using this function requests DW to create a local table, comprising molecular structure, compound name, Hill formula, and molecular weight for each entry about chemical compounds in the (English
version of) wikipedia. The source for this is the little «info box» the English wikipedia shows on the right hand top corner of each article. A priori both molecular [organic] compounds (like lutidine) as well as inorganic compounds (like sodium chloride) are considered; currently summing up 17421 entries.

This local table may serve you as starting point for any computations with DW including the search
by name, (sub)structure and structure similarity; a click with the right mouse button will direct you to the corresponding wikipedia entry. The functionality is related to the on-line service of Wikipedia Chemical Structure Explorer at http://www.cheminfo.org/wikipedia.

Its «inner working» was published in

Ertl, P., Patiny, L., Sander, T., Ruferner, C., Zasso, M. Wikipedia Chemical Structure Explorer: substructure and similarity searching of molecules from Wikipedia. J Cheminform 7, 10 (2015)

as open access article https://doi.org/10.1186/s13321-015-0061-y