Subject: Exporting a descriptor as a Textfile
Posted by Christophe on Mon, 09 May 2022 15:45:56 GMT
View Forum Message <> Reply to Message

Hello everyone,

Is it possible to export descriptors such as Skelspheres or others as a textfile ?

Thanks

Subject: Re: Exporting a descriptor as a Textfile
Posted by Christophe on Wed, 18 May 2022 12:22:27 GMT
View Forum Message <> Reply to Message

Just to mention that I'd like to use these text files as a part of a matrix to perform diverse ordination techniques not available in DW like UMAP for example.
These files must be stored somewhere since the "Skelsphere" descriptor for example is available when a t-SNE ordination is envisioned.

Thanks

Subject: Re: Exporting a descriptor as a Textfile
Posted by amorrison on Thu, 19 May 2022 10:42:52 GMT
View Forum Message <> Reply to Message

Hi,
If you use 'add calculated values' and then str(Descriptor Variable) a new column should appear.
I think this is what you are looking for.
Angus

Subject: Re: Exporting a descriptor as a Textfile
Posted by Christophe on Fri, 20 May 2022 14:23:56 GMT
View Forum Message <> Reply to Message

Hi Angus,

Thank you
The columns appear well but the data are unusable (cabalistic signs)
Some descriptors are encoded as a matrix (1024 bits for example), so I suppose this is not the proper way to extract it.

Christophe

Christophe,

for a small set of molecules, I think I'm able to replicate your findings e.g., for the assignment and subsequent display of skelspheres as a string:

(observation with DW 5.5.0 for Linux, including the update by May 13th).

On the other hand (now specific to skelspheres); if these are like fingerprints e.g., openbabel offers,[1] what are programs you intend to use which accept these as an input for further computation?

E.g., openbabel reports about DMF:

```
$ obabel -:"CN(C)C=O" -ofpt -xf FP2
>   5 bits set
00000000 00000000 00000000 00000000 00000000 00000000
00000000 00000000 00000000 00000000 00200000 00000000
00008000 00000000 00000000 00000000 00000000 00000000
00000000 00000000 00000000 00000400 00000000 00000000
00000000 00000000 00000000 00000000 00000000 00040001
00000000 00000000
1 molecule converted
```

or

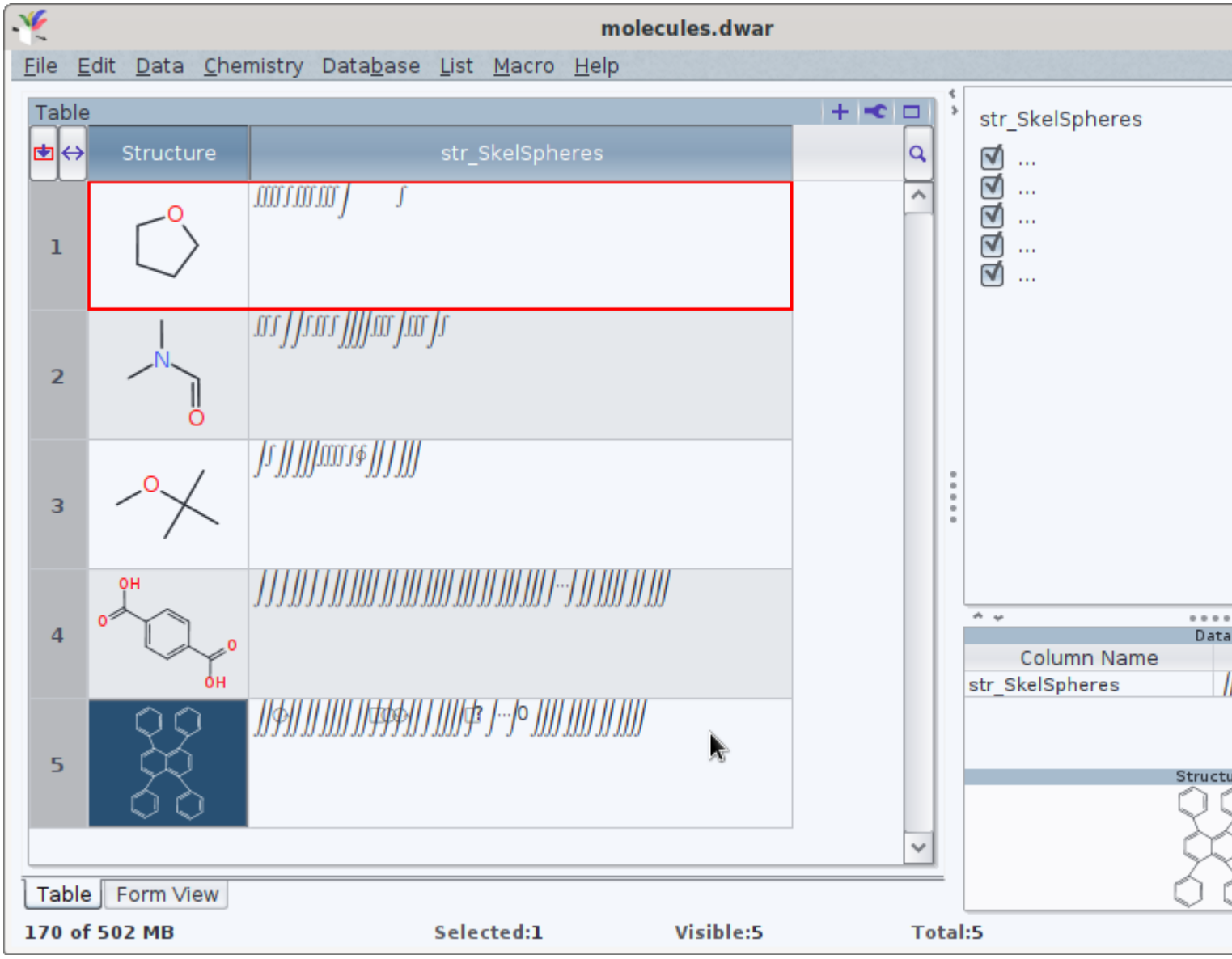
```
$ obabel -:"CN(C)C=O" -ofpt -xs -xf FP2
>
0 6 1 7 1 6 <693>
0 7 1 6 <82>
0 8 2 6 <623>
0 8 2 6 1 7 <330>
0 8 2 6 1 7 1 6 <64>
1 molecule converted
```

Norwid

[1] https://open-babel.readthedocs.io/en/latest/FileFormats/Fing erprint_format.html

File Attachments

1) example_str_skelspheres.png, downloaded 1585 times



2) molecules.dwar, downloaded 633 times

Subject: Re: Exporting a descriptor as a Textfile
Posted by Christophe on Tue, 24 May 2022 13:24:35 GMT
View Forum Message <> Reply to Message

Hi Norwid,

Thank you

It looks you managed to convert the structure DW column into one of the finger print managed by open babel.

I am not sure this procedure captures the native information of the skelspheres descriptor. It generated a FP2 (by default) FP descriptor.

I know how to generate a lot of different FP from structures. For example PaDEL (free java program, http://www.yapcwsoft.com/dd/padeldescriptor) does a very good job. It provides a matrix (.csv file) with structures (in rows) and bits (binary or count, depending on the FP you select) (in columns).

If you know how bits (or series of bits) are organized you can try to trace the source that causes the differences in distribution by multidimension reduction methods in R for example.

with DW, when I apply a similarity (or activity cliff) with Skelspheres and/or OrgFunctions, I have data sets that cluterize very well but it is quite challenging then to relate the resulting clusters to the distributional differences in term of structure (SkelSpheres) or functionalization (Orgfunctions). If these two descriptors are different from the one I used so far such as ExtFP, MACCSFP .... may be I could gain more information. But I need the "matrix formalism" into a text file.
Christophe

Subject: Re: Exporting a descriptor as a Textfile
Posted by thomas on Thu, 02 Jun 2022 07:46:22 GMT
View Forum Message <> Reply to Message

Hi Christoph,

if you use Java, you could use this line to decode the SkeletonSpheres descriptor into a byte array, which contains 1024 count values:

byte[] counts = new DescriptorHandlerSkeletonSpheres().decode(encodedSkeletonSph eres);

Then you could loop over the counts array and write numbers where ever you want. The only dependency would be OpenChemLib, which you can find on GitHub.

Likewise you can decode the OrgFunctions descriptor with

int[][] pairs = new DescriptorHandlerFunctionalGroups().decode();

Here you get an array of arrays with length of 2. Every one of these small arrays contains a functional group ID and an associated count value. Thus, this is not a simple matrix and making use of it will probably need the some knowledge of the groups, i.e. the similarity tree. You may study the FunctionalGroupClassifier to understand which groups have which ID and how the tree is organized.

By the way, UMAP support in DataWarrior is planned.

Thomas

Subject: Re: Exporting a descriptor as a Textfile
Posted by Christophe on Thu, 02 Jun 2022 15:24:52 GMT
View Forum Message <> Reply to Message

Hello Thomas,

Thank you for this valuable information as usual.

UMAP support in future versions of DW is a very good news. Great!!!

Thanks

Christophe

Subject: Re: Exporting a descriptor as a Textfile
Posted by thomas on Wed, 13 Jul 2022 15:29:00 GMT
View Forum Message <> Reply to Message

Hi Christophe,

UMAP is now available...

Subject: Re: Exporting a descriptor as a Textfile
Posted by Christophe on Mon, 18 Jul 2022 11:35:14 GMT
View Forum Message <> Reply to Message

Hello Thomas,

Thanks for this news.

But we are still at the 5.50 version aren't we ?

When I check for updates, the system tells me that I've got the latest one, i.e. v5.50.

When did you plan to release the latest version?

Christophe

Subject: Re: Exporting a descriptor as a Textfile
Posted by nbehrnd on Tue, 19 Jul 2022 05:54:15 GMT
View Forum Message <> Reply to Message

Christophe,

have a look at https://openmolecules.org/datawarrior/download.html. Click on «I have read and understood the disclaimer» and you find the links to the archives for Linux/Mac, or for Windows to obtain the updates (i.e., you need an already functional installation in first place). Select the one suitable for you, decompress the archives, and substitute with the .jar they include the .jar of your installation.

Norwid

For Mac and Linux: https://openmolecules.org/datawarrior/dw550x.zip
For Windows: https://openmolecules.org/datawarrior/dw550win.zip

---

Subject: Re: Exporting a descriptor as a Textfile
Posted by Christophe on Tue, 19 Jul 2022 07:32:27 GMT
View Forum Message <> Reply to Message

Hi Norwid,

Thank you.

Indeed, the whole procedure is explained on the website.
I hadn't even bothered to read it.

Christophe
