Subject: suggest: native .pdf export
Posted by nbehrnd on Mon, 11 May 2020 10:29:36 GMT
View Forum Message <> Reply to Message

Based on a .smi derived from elsewhere,[1] which was converted with openbabel into a .sdf I let DataWarrior screen the structures for a few characteristics.

Aiming to share it with one who does not use DataWarrior, I miss a built-in function to print the table as a .pdf file as a vehicle of discussion.
This could be useful, especially if -- as here -- the cell's background colour has a significance.

So far, I printed the table via an installed HP printer as a postscript into a file (of very large file size) and converted it with ps2pdf into a .pdf like the example file attached. At page breaks, however, lines seem to be broken (table headings). If exported directly from DW as .pdf, their file size could be considerably smaller than now (containing an image in the .pdf as a container) and possibly retain a searchable text-layer.

[1] https://pubs.acs.org/doi/suppl/10.1021/jm301008n/suppl_file/ jm301008n_si_002.xlsx

File Attachments
1) table_S3.png, downloaded 1384 times
2) table_S3.dwar, downloaded 653 times
3) table_S3.pdf, downloaded 689 times

Subject: Re: suggest: native .pdf export
Posted by thomas on Mon, 11 May 2020 18:51:53 GMT
View Forum Message <> Reply to Message

If you have a pdf printer driver installed (on Ubuntu you do that with "sudo apt-get install printer-driver-cups-pdf"),you can easily print as PDF, see attached file.

A comment to your openbabel conversion: DataWarrior can directly interpret SMILES. If you rename your .smi to .txt and open with DataWarrior, you should automatically get a structure column. An alternative is to copy the SMILES into the clipboard and just paste them into DataWarrior, which produces a new document with native chemical structures.

Best wishes, Thomas

File Attachments
1) DataWarrior_table_S3-1.dwar__on-Linux-generated_files-job_91.pdf, downloaded 675 times

Subject: Re: suggest: native .pdf export
Posted by nbehrnd on Tue, 12 May 2020 14:03:44 GMT
View Forum Message <> Reply to Message

It was possible to replicate the indicated method using the cups pdf printer.

As a closing comment:
Still interested to benefit more from the vector format I wrote a Python script
that reads some of DW's .dwar file content and the retained list of SMILES strings,
calls openabel to visualize the structures, and puts all in an .xlsx file. The
manual work then left was to open this file in LibreOffice Calc, to adjust the
images' sizes to fit the cell size, to apply conditional cell background colors
and to save it as .ods. The file size of the then exported .pdf is slightly less
than half of the one printing from DW with cups while still offering a searchable,
crisply printed text layer, too.

## File Attachments

1) table_S3.smi, downloaded 600 times
2) table_S3.dwar, downloaded 637 times
3) spreadsheet_test.py, downloaded 712 times
4) test.ods, downloaded 641 times
5) test.pdf, downloaded 709 times

Subject: Re: suggest: native .pdf export
Posted by nbehrnd on Fri, 15 May 2020 11:11:21 GMT
View Forum Message <> Reply to Message

There is a reason why the structure import used a .sdf generated from the .smi by openbabel,
instead of reading the SMILES listing file directly.
Neither the direct read of the smiles from a file like 3entries.smi.txt, nor the special
copy / paste from the clipboard (paste without header) kept the column about molecules'
names with e.g., the PubChem number only. The attached .pdf documents this observation.

If file 3entries.smi.txt containing the annotated smiles starts with an explicit header
line, e.g "structure SMILES", then DW reads the file as containing three structures; the
annotating column then still contains both SMILES string and the PubChem number entry.

## File Attachments

1) SMILES_3entries.smi.txt, downloaded 557 times
2) 3entries.sdf, downloaded 639 times
3) 3entries_import_sdf.dwar, downloaded 599 times
4) 3entries_import_smi.dwar, downloaded 594 times
5) data_read.pdf, downloaded 736 times

Subject: Re: suggest: native .pdf export
Posted by thomas on Thu, 21 May 2020 22:25:28 GMT
View Forum Message <> Reply to Message

DataWarrior uses TAB or comma delimited text files. The smi File only contains spaces between Smiles and Name. If you replace all SPACEs by a TABs in any Text-Editor before pasting into DataWarrior, you will correctly get three columns: Structure, Smiles, and Name