

OSRA: GPL-ed molecule drawing to SMILES convertor

Egon Willighagen 

Published July 20, 2007

Citation

Willighagen, E. (2007). OSRA: GPL-ed molecule drawing to SMILES convertor. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/arc2j-1ha32>

Keywords

Cheminf, Openbabel

Abstract

Igor wrote a message to the CCL mailing list about OSRA:

Copyright

Copyright © Egon Willighagen 2007. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

Igor wrote a message to the [CCL mailing list](#) about [OSRA](#):

We would like to announce a new addition to the set of cheminformatics tools available from the Computer-Aided Drug Design Group at the NCI-Frederick. OSRA is a utility designed to convert graphical representations of chemical structures, such as they appear in journal articles, patent documents, textbooks, trade magazines etc., into SMILES.

|

OSRA can read a document in any of the over 90 graphical formats parseable by ImageMagick (GIF, JPEG, PNG, TIFF, PDF, PS etc.) and generate the SMILES representation of the molecular structure images encountered within that document.

The email does not give any information on the fail rate, but the demo they provide via the [webinterface](#) does show some minor glitches (the bromine is not recognized):

The screenshot displays the OSRA web interface. On the left, under the 'Upload Image File' tab, a chemical structure image is shown, labeled '8 of 9'. The structure is a complex molecule with a bromine atom, a nitrile group, and a phosphonate group. On the right, under the 'Edit Structure' tab, the same chemical structure is displayed in a drawing tool. The drawing tool has a toolbar with various icons for editing and a vertical menu on the left with letters C, N, O, S, F, Cl, Br, I, P, X. The bromine atom in the structure is highlighted in pink, and a question mark is visible below the nitrile group, indicating a recognition error.

The source reuses [OpenBabel](#) and uses the GPL license. The value equal to that of text mining tools like [OSCAR3](#), and together they sounds like the Jordan and Phippen of mining chemical literature.