

Optical Chemical Structure Recognition

Egon Willighagen 

Published July 31, 2007

Citation

Willighagen, E. (2007). Optical Chemical Structure Recognition. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/gh6e5-t2g74>

Keywords

Cheminf

Abstract

Days after the release of OSRA last week, I saw the optical chemistry structure recognition on the front page of my favorite Dutch /. equivalent, Tweakers.net, Duitsers leren computer chemische structuren herkennen, written by René Gerritsen.

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.

Days after the release of [OSRA](#) last week, I saw the *optical chemistry structure recognition* on the front page of my favorite Dutch /. equivalent, [Tweakers.net](#), [Duitsers leren computer chemische structuren herkennen](#), written by [René Gerritsen](#). The article discusses the Fraunhofer Institute's [ChemoCR](#), which was, IIRC, presented as poster at last year's [German Conference on Chemoinformatics](#) (to be held again [this year](#)). Meanwhile, the CCL.net mailing list had a [discussion on the alternatives](#) too; I think it is fair to say that the chemical community realizes the importance of these tools. Below is a short overview of the available tools, including some important information regarding integration into workflows.

ChemoCR

ChemoCR seems to be proprietary software, as I could not find any download, and InfoChem seems to be the party to sell licenses. The [screenshot](#) in the Tweakers.net article seems to show that it is written in Java, but that hardly matters if not open source. The project is said to have started three years ago.

CLiDE

[CLiDE](#) is another commercial (expensive) program to do the job. It was developed more than ten years ago, and the [most recent scientific publication](#) is from 1997 (as the webpage states).

OSRA

[OSRA](#) (see my [previous blog](#)) is opensource and uses the GPL license. It is written in C++. It does not as feature complete as ChemoCR yet, but that will surely come. This project is surely the youngest project.

Kekule

I have not picked up copy of the paper [Kekule: OCR-optical chemical \(structure\) recognition](#) cited by [Tony](#), so cannot say much about that right now.

It is obvious that only OSRA lends itself to embedding in reproducible workflows. Debra Banville [reviewed the two commercial programs CLiDE and ChemoCR](#) last year, along with a few other text mining tools in chemoinformatics. I am curious about her opinion of the new opensource tools in this arena.