

Bioclipse-Oscar4 - Text mining in Bioclipse

Egon Willighagen 

Published September 27, 2011

Citation

Willighagen, E. (2011, September 27). Bioclipse-Oscar4 - Text mining in Bioclipse. *Chem-bla-ics*.
<https://doi.org/10.59350/qgrq1-4r761>

Keywords

Oscar, Bioclipse, Beilstein

Abstract

Almost a year ago I started a position with Peter Murray-Rust to work on Oscar for three months (see this overview of results; a paper by the full Oscar team (Sam, David, Dan, Lezan) is pending, and I'm really happy to have been able to contribute bits to the project). Since then, I have had little time :(That's how it goes, with post-hopping, unfortunately. One thing I did do after that, was write a Bioclipse plugin.

Copyright

Copyright © Egon Willighagen 2011. 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.

Almost a year ago I [started a position](#) with [Peter Murray-Rust](#) to work on Oscar for three months (see this overview of results; a paper by the full Oscar team (Sam, David, Dan, Lezan) is pending, and I'm really happy to have been able to contribute bits to the project). Since then, I have had little time :(That's how it goes, with post-hopping, unfortunately. One thing I did do after that, was write a [Bioclipse plugin](#).

I was asked recently via [LinkedIn](#) if I was planning a Bioclipse-Oscar plugin, and I realized that I forgot to blog about it. So, here goes. The **oscar** manager I implemented follows the [Oscar API](#) , and these methods are available: `extractText()`, `findNamedEntities()`, `findResolvedNamedEntities()`.

When I wrote the plugin, I also uploaded an [example workflow to MyExperiment](#). The code is:

```
// Demo showing the Oscar text mining functionality
// in Bioclipse
var html = bioclipse.download(
    "http://dx.doi.org/10.3762/bjoc.6.133",
    "text/html"
)
var text = oscar.extractText(html);
// the next step may take some time, while
// initializing the Oscar software for the
// first time
var mols = oscar.findResolvedNamedEntities(text);
var file = "/Oscar Demo/extractedMols.sdf";
cdk.saveSDFFile(file, mols);
ui.open(file);
```

The code will extract chemical entities, and open a molecules table in [Bioclipse](#):

The screenshot displays the Bioclipse application window. The interface is divided into several panels:

- Left Panel (Project Explorer):** Shows a hierarchical tree of project folders including BC - OpenTox, ChemBioMod, CHEMINF Paper, FooQSAR, Gists, Molecules, NanoTox, OpenTox, Oscar Demo, OSDD, PharmBio, PharmBio Data, QSAR, Sample Data, Temp, Test, ToxBank, and Virtual. The 'Oscar Demo' folder is expanded, showing files like extractedMols.sdf, oscar2OpenTox.js, oscarDemo.js, and oscarDemo.png.
- Top Panel (Code Editor):** Displays the file 'oscarDemo.js' with the following JavaScript code:

```
1 // Demo showing the Oscar text m
2
3 var html = bioclipse.download(
4   "http://dx.doi.org/10.3762/bjo
5   "text/html"
6 )
7
8 var text = oscar.extractText(htm
9
10 // the next step may take some t
11 // software for the first time
12 var mols = oscar.findResolvedNam
13
14 var file = "/Oscar Demo/extracte
15 cdk.saveSDFFile(file, mols);
16 ui.open(file);
17
```
- Right Panel (Molecule Viewer):** Displays a table with two rows of chemical structures. The first row, labeled '93', shows a 2D structure of a molecule with a blue nitrogen atom, a red oxygen atom, and two black carbon atoms. The second row, labeled '94', shows a 2D structure of a benzene ring. The table has tabs for 'Table', 'Single Molecule', and 'Headers'.
- Bottom Panel (JavaScript Console):** Shows the command prompt output for the 'man oscar' command:

```
> man oscar
-----
oscar
-----
Adds text mining functionality to Bioclipse.

This manager has the following methods:
oscar.extractText( String html )
oscar.findNamedEntities( String text )
oscar.findResolvedNamedEntities( String text )
```

The status bar at the bottom indicates '266M of 440M' memory usage.