chem-bla-ics

Oscar4 paper: text mining in Bioclipse (and everywhere else, of course)

Egon Willighagen

Published November 1, 2011

Citation

Willighagen, E. (2011). Oscar4 paper: text mining in Bioclipse (and everywhere else, of course). In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/rz0tz-3wa91

Keywords

Oscar, Bioclipse, Myexperiment

The Oscar4 paper (CC-BY, just like the screenshots of the paper below) was out already some days now, but the formatting has finished:

Jessop et al. Journal of Cheminformatics 2011, 3:41
http://www.jcheminf.com/content/3/1/41

Journal of Cheminformatics

SOFTWARE Open Access

OSCAR4: a flexible architecture for chemical text-mining

David M Jessop, Sam E Adams, Egon L Willighagen, Lezan Hawizy and Peter Murray-Rust*

Abstract

The Open-Source Chemistry Analysis Routines (OSCAR) software, a toolkit for the recognition of named entities and data in chemistry publications, has been developed since 2002. Recent work has resulted in the separation of the core OSCAR functionality and its release as the OSCAR4 library. This library features a modular API (based on reduction of surface coupling) that permits client programmers to easily incorporate it into external applications. OSCAR4 offers a domain-independent architecture upon which chemistry specific text-mining tools can be built, and its development and usage are discussed.

Introduction

In keeping with the historical and methodological aspects of this special issue, we recount the history and motivation of OSCAR.

A large amount of factual data in chemistry and neighbouring disciplines is published in the form of text [14,15], represent the public state-of-the-art in chemical text analysis and extraction.

The OSCAR (Open-Source Chemistry Analysis Routines) software has been developed over a period of years and a number of projects. Between 2002 and 2004, sponsors including the Royal Society of Chemistry

I spotted a rogue `http://` in the code example b) in Appendix B:

b) Where the named entity can be resolved to a chemical structure, extract it:

```
Oscar oscar = newOscar();
List < ResolvedNamedEntity >entities
= oscar.findAndResolveNamedEntities(s);
for (ResolvedNamedEntity entity : entities) {
ChemicalStructure structure = entity.get-
FirstChemicalStructurehttp://(FormatType.
INCHI));
...
}
```

I'll see what I can do about that, but the API might evolve a bit anyway.

That leaves me to mention that Bioclipse has an Oscar extension (Bioclipse has a lot of functionality nowadays, in fact), and that I blogged several times on Oscar4 when I was working with the other authors on the refactoring last year.