Chemistry Central Journal

Poster presentation

Creating chemo- and bioinformatics workflows, further developments within the CDK-Taverna Project

T Kuhn*[1,2], A Zielesny[2] and C Steinbeck[3]

Address: [1]Cologne University Bioinformatics Center (CUBIC), Cologne, Germany, [2]University of Applied Sciences of Gelsenkirchen, Institute for Bioinformatics and Chemoinformatics, Recklinghausen, Germany and [3]European Bioinformatics Institute (EBI), Cambridge, UK

* Corresponding author

from 4th German Conference on Chemoinformatics
Goslar, Germany. 9–11 November 2008

Published: 5 June 2009

Chemistry Central Journal 2009, **3**(Suppl 1):P42 doi:10.1186/1752-153X-3-S1-P42

This abstract is available from: http://www.journal.chemistrycentral.com/content/3/S1/P42

The CDK-Taverna project aims at building an open-source pipelining solution through combination of different open-source projects such as Taverna [1] the Chemistry Development Kit (CDK) [2] and Bioclipse [3].

Pipelining or workflow tools allow for the Lego™-like, graphical assembly of I/O modules and algorithms into a complex workflow which can be easily deployed, modified and tested without the hassle of implementing it into a monolithic application. Current developments in CDK-Taverna focus on a soft computing framework which allows a flexible use of different methods from, for example, the WEKA [4] library. Additionally an implementation of the ART2a [5] algorithm is added as worker and could be used for cluster purposes. Worker which analyse and compare the cluster results are also implemented. Pgchem::Tigress [6] a chemoinformatic cartridge for a PostgresSQL database is used to allow the handling of large datasets. Therefore a number of worker for adding, selection and searching molecules from and on the database are added.

References

1. Oinn T, Addis M, Ferris M, Marvin D, Senger M, Greenwood M, Carver T, Glover K, Pocock M, Wipat A, Li P: **Taverna: A tool for the composition and enactment of bioinformatics workflows.** *Bioinformatics* 2004, **20(17):**3045-3054.
2. Steinbeck C, Han YQ, Kuhn S, Horlacher O, Luttmann E, Willighagen E: **The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics.** *J Chem Inf Comput Sci* 2003, **43:**493-500.
3. Spjuth O, Helmus T, Willighagen EL, Kuhn S, Eklund M, *et al.*: *Bioclipse: An open rich client workbench for chemo-and bioinformatics* in press.
4. Witten IH, Frank E: **Data Mining: Practical machine learning tools and techniques.** 2nd edition. *Morgen Kaufmann, San Francisco*; 2005.
5. Carpenter GA, Grossberg S, Rosen DB: *Neural Networks* 1991, **4:**493.
6. Schmid EG: [http://pgfoundry.org/projects/pgchem/].