Chemistry Central Journal

Poster presentation

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

Thomas Kuhn*[1,2], Achim Zielesny[1] and Christoph Steinbeck[1]

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

* Corresponding author

from 3rd German Conference on Chemoinformatics
Goslar, Germany. 11-13 November 2007

Published: 26 March 2008

Chemistry Central Journal 2008, **2**(Suppl 1):P27 doi:10.1186/1752-153X-2-S1-P27

This abstract is available from: http://www.journal.chemistrycentral.com/content/2/S1/P27

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. Here, properties of chemical substances may be calculated using descriptors from the QSAR / QSPR package of the Chemistry Development Kit (CDK).

Further, a reaction enumeration algorithm for combinatorial chemistry based on existing methods of the Chemistry Development Kit is being developed. This algorithm allows for the enumeration of a reaction given that reactants and products are provided as "Markush" structures.

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 V, *et al.*: **An open rich client workbench for chemo- and bioinformatics.** . *submitted.*
4. Witten IH, Frank E: **Data-Mining Practical machine learning tools and techniques.** 2nd Edition edition. *Morgen Kaufmann, San Francisco*; 2005.
5. Hassan M, Brown RB, Varma-O'Brien , Rogers D: **Cheminformatics analysis and learning in a data pipelining environment.** *Molecular Diversity* 2006, **10:**283-299.