

POSTER PRESENTATION

Open Access

Structured chemical class definitions and automated matching for chemical ontology evolution

Lian Duan^{1,2*}, Janna Hastings^{1,3}, Paula de Matos¹, Marcus Ennis¹, Christoph Steinbeck¹

From 7th German Conference on Chemoinformatics: 25 CIC-Workshop
Goslar, Germany. 6-8 November 2011

Ontologies encode the knowledge of human experts in order to allow computers to automate common tasks in a domain. They are hierarchically organised and backed by computational logic which allows automated inferences of the implicit consequences of explicitly stated knowledge. ChEBI is a database and ontology of chemical entities of biological interest [1]. Within the ontology, chemical entities are classified based on shared structural features and also based on their roles and activities in biological systems. For example, the chemical class 'aminopyridine' is defined as 'Compounds containing a pyridine skeleton substituted by one or more amine groups', while an example of a role based class is 'antiviral drug', which groups together chemical entities that are used as antiviral drugs, regardless of their chemical structure. We have developed a novel semi-automated system for creating structure-based chemical class definitions. Our tool allows curators to draw and visually define shared structural features for classes of chemicals, which definitions are then used to automatically detect class membership across the full chemical database. The front end is based on an extended JChemPaint [2] and the Google Web Toolkit, and the back-end on a custom extension of the Chemistry Development Kit [3]. With this tool, it is possible to define chemical classes based on molecular skeletons, substitute groups, arbitrary parts including cycles of arbitrary length, formulae and overall properties, and these features can be combined using nested logical operators. Matching these definitions to candidate structures from the database is accomplished by means of an in-memory matching procedure, validated against the existing manually

curated classification in ChEBI, allowing us to iteratively refine both the definitions of classes as well as to evolve the quality of the classification in ChEBI.

Author details

¹European Bioinformatics Institute, Cambridge, UK. ²ETH, Zurich, Switzerland. ³University of Geneva, Switzerland.

Published: 1 May 2012

References

1. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburner M: **ChEBI: a database and ontology for chemical entities of biological interest.** *Nucl Acids Res* 2008, **36**(Suppl. 1):D344-D350.
2. Steinbeck C, Han Y, 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. Krause S, Willighagen E, Steinbeck C: **JChemPaint- Using the Collaborative Forces of the Internet to Develop a Free Editor for 2D Chemical Structures.** *Molecules* 2000, **5**(1):93-98.

doi:10.1186/1758-2946-4-S1-P5

Cite this article as: Duan et al.: Structured chemical class definitions and automated matching for chemical ontology evolution. *Journal of Cheminformatics* 2012 **4**(Suppl 1):P5.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
<http://www.chemistrycentral.com/manuscript/>


ChemistryCentral

* Correspondence: dlian@ebi.ac.uk

¹European Bioinformatics Institute, Cambridge, UK

Full list of author information is available at the end of the article