CDK Literature #4

Citation

Keywords

Cdk

Abstract

Fourth in the CDK Literature series. Really, a follow up on #3 which I wanted to get out, even though not really finished yet. But, after 3 comes 4, not 3b. Maybe 3.1, but that suggests at least 3.2-3.9 too, let alone full R (that was supposed to the space of all reals...) I'll stick to positive non-zero integers. #1 and #2 are still available too.

Copyright

chem-bla-ics

Fourth in the *CDK Literature* series. Really, a follow up on #3 which I wanted to get out, even though not really finished yet. But, after 3 comes 4, not 3b. Maybe 3.1, but that suggests at least 3.2-3.9 too, let alone full R (that was supposed to the space of all reals...) I'll stick to positive non-zero integers. #1 and #2 are still available too.

Another thing I should remark is that this series does not provide full reviews of the cited papers. Instead, it provides a list of papers that cite one of the two CDK papers (doi:10.1021/ci025584y and doi:10.2174/138161206777585274). It is worth repeating that these two articles are not the only article which describe CDK source code, and maybe I should start listing papers that cite other articles that discuss CDK source code. Anyway, the other papers discussing CDK source code are listed in an overview I maintain in the CDK wiki (now on SourceForge), but have not updated it for #4 and #3 yet.

Organic Reaction Ontology

Punnaivanam Sankar and Gnanasekaran Aghila published a paper where they propose a knowledge framework for mechanisms of organic reactions, and used an XML framework combined with a ontology for the semantics. JChemPaint and the CDK are cited as opensource tools that support reactions.

Punnaivanam Sankar, Gnanasekaran Aghila, Ontology Aided Modeling of Organic Reaction Mechanisms with Flexible and Fragment Based XML Markup Procedures, J. Chem. Inf. Model., 2007, 47(5):1747 -1762, doi:10.1021/ci700043u

More QSAR

Ma et al. have published a (Chinese) QSAR paper where CDK descriptors have been as molecular represention of a data set with 212 ligands for the P-glycoprotein. Models have been build with Random Forests, and classification success rates for the test set of around 85%.

Guang-Li Ma, Xiao-Ping Zhao, Yi-Yu Cheng, Identification of P-gp substrates using a random forest method based on chemistry development kit descriptors, Chemical J. of Chinese Universities-Chinese, 2007, 28(10):1885-1888

Chemical Databases

sMOL is GPL-licensed software for setting up a small molecule database. The software uses JChemPaint, OpenBabel, JOELib and the CDK for chemoinformatics functionality, and R and Weka for statistical analyses. I have not locally installed it yet, but the User Guide shows really nice screenshots. The Installer Guide shows a quite polished product too. Not sure how open the project is to contributions from others (patches, translations, etc, but will ask.

Supawadee Ingsriswang, Eakasit Pacharawongsakda, sMOL Explorer: an open source, web-enabled database and exploration tool for Small MOLecules datasets, Bioinformatics, 2007, 23(18):2498-2500, doi:10.1093/bioinformatics/btm363

Free Tools

Bruno Villoutreix wrote an overview of free (as in free beer) services to aid virtual screening. It cites the CDK, Jmol, OpenBabel as tools, along with a long list of free but proprietary tools. It does explicitly plead for opensource docking and scoring tools, and, as such, potentially useful in grant proposals.

Bruno Villoutreix, Nicolas Renault, David Lagorce, Olivier Sperandio, Matthieu Montes, Maria Miteva, Free Resources to Assist Structure-Based Virtual Ligand Screening Experiments, Current Protein and Peptide Science, 2007, 8(4):381-411, doi:10.2174/138920307781369391

Metabolomics

Fangping Mu et al. have set up a KEGG-derived database with annotated reactions where atoms between reactants and products are mapped, to help data analysis of isotopomeromics data. The CDK rendering features are used for visualization purposes. The software also builds on BioMeta, work by Martin Ott , presented at last years CDK Workshop.

Fangping Mu, Robert Williams, Clifford Unkefer, Pat Unkefer, James Faeder, William Hlavacek, Carbon-fate maps for metabolic reactions, Bioinformatics, 2007, 23(23):3193-3199, doi:10.1093/bioinformatics/btm498

I got two more papers lined up, but do not have access to Current Pharmaceutical Design.