

ter a discussion on the `cdk-devel@lists.sf.net` mailing list: <http://qsar.sf.net/>. This project aims to bring together open source developers from many projects and develop a Java GUI program that interfaces with all aspects of QSAR(-like) research: setting up a data set, descriptor calculation, model building, up to model validation.

The CDK project is expected to contribute to this project by providing implementations of several of these components.

SMARTS

Recently, the `UniversalIsomorphismTester` was adapted to allow for custom `Atom-Atom` and `Bond-Bond` matching. Prior to this change `Atoms` were matched based only on element symbol. As a result it was not possible to distinguish an sp^2 and sp^3 carbon. In addition, it was not possible to match an atom to any halogen. This shortcoming has been fixed now.

The next step is to write a SMARTS [4] parser and editor that can create `SMILESAtoms` that can match real atoms based on the given query. This subproject has been started recently, but the full query language is not implemented yet. A basic example has been implemented (see `cdk.test.isomorphism.SMARTSTest`). In this example the SMARTS query `'C=*` is used, thus a carbon double bonded to any atom. This is the source code that implements this:

```
SmilesParser sp = new SmilesParser();
AtomContainer atomContainer = sp.
    parseSmiles("CC(=O)OC(=O)C");
// acetic acid anhydride
QueryAtomContainer query =
    new QueryAtomContainer();
SMARTSAtom atom1 = new SMARTSAtom();
atom1.setLabel("*");
SMARTSAtom atom2 = new SMARTSAtom();
atom2.setSymbol("C");
query.addAtom(atom1);
query.addAtom(atom2);
```

Literature

"Literature" is a recurrent column describing recently published articles that have in some way to do with CDK.

by Egon Willighagen

This column intends to give an overview of recently published articles that have some relation to CDK: they might describe algorithms implemented

```
query.addBond(
    new OrderQueryBond(atom1, atom2, 2)
);
boolean isSubstructure =
    UniversalIsomorphismTester.
    isSubgraph(atomContainer, query);
```

The `SMARTSAtom.match()` method only implements the `*` atom and much needs to be done before this fully works. Feel free to browse the source code in the `cdk.smiles.smarts` package.

More

These three things are not the only ongoing development of CDK, but show three very interesting new features. People are encouraged to read the *CDK ChangeLog* which will appear in each issue. But here are some keywords: more reactions, tighter CML support, partial atomic charges and more CDK plugins.

Egon Willighagen
University of Nijmegen, The Netherlands
egonw@sci.kun.nl

Bibliography

- [1] M.L Allinger. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms. *J. Am. Chem. Soc.*, 99, 1977.
- [2] Ghemical. <http://ghemical.sf.net/>, April 2004.
- [3] R. Todeschini and V. Consonni. *The Handbook of Molecular Descriptors*, volume 11 of *Methods and Principles in Medicinal Chemistry*. Wiley-VCH, Weinheim, Germany, 2000.
- [4] Daylight website. <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>, April 2004.

in CDK, use CDK in research, or describe software that uses CDK. Normally, this article will discuss articles published since the previous issue, but since this is the first issue, it will describe all CDK related publications that appeared so far.

The articles will be described in the order in which they appeared, but I'll take the liberty to start with the CDK article itself.

The CDK article

Early 2003, the CDK article was published [?]. It describes the CDK project, and an important part of the architecture of the library. A must read :). Therefore, I will not further discuss it here.

JChemPaint

JChemPaint's publication appeared in 2000 in a special issue of *Molecules* <http://www.mdpi.net/molecules/> dedicated to a virtual conference where a poster featured the program [1]. The article does not describe the CDK based JChemPaint version which was released around the same time this issue was published, but describes the older program. Note that some algorithms in CDK actually originate from this JChemPaint version, or one of its libraries.

CML Reading

The CML reading algorithm that is used by the `cdk.io.CMLReader` was originally written for Jmol and JChemPaint and was originally published on the Chemistry Preprint Server [2], and in 2001 appeared in the *Internet Journal of Chemistry* [3]. The article describes the SAX based XML reading and describes how it deals with CML conventions.

NMRShiftDB

Last year, another CDK-based piece of chemoinformatics software, NMRShiftDB, was released to the academic community. Technical details have been published [4]. The database is described in a separate article in this newsletter.

Chemistry enriched RSS

An extension of Rich Site Summary (RSS) [5] with chemical information was introduced last January coined CMLRSS [6]. RSS is a system used by websites to distribute news headlines over the internet.

Currently, the headlines only contain a simple textual description, though embedding bibliographic information has been proposed by some scientific journals. CMLRSS extends RSS by embedding molecular information in CML format. The article describes a CDK plugin, the RSSViewer, which is able to download news feeds and extract chemistry (crystal/molecular structures) and display that in Jmol, JChemPaint or any other CDK plugin aware program.

Though I'll try to keep up with literature, I might oversee an interesting article that mentions, uses, extends, compares or otherwise relates to the CDK. To make sure it gets covered in this column you can send me an email with the bibliographic information for that article.

Egon Willighagen
University of Nijmegen, The Netherlands
egonw@sci.kun.nl

Bibliography

- [1] Stefan Krause, Egon Willighagen, and Christoph Steinbeck. JChemPaint - Using the Collaborative Forces of the Internet to Develop a Free Editor for 2D Chemical Structures. *Molecules*, 5:93–98, 2000.
- [2] Chemical Preprint Server. <http://www.chemweb.com/preprint>.
- [3] E.L. Willighagen. Processing CML Conventions in Java. *Internet Journal of Chemistry*, 4, 2001.
- [4] C. Steinbeck, S. Kuhn, and S. Krause. NMR-ShiftDB - Constructing a Chemical Information System with Open Source Components. *J. Chem. Inf. Comput. Sci.*, 43(6):1733 – 1739, 2003.
- [5] RDF Site Summary 1.0 (RSS). <http://web.resource.org/rss/1.0/spec>.
- [6] Mark J. Williamson Peter Murray-Rust, Henry S. Rzepa and Egon L. Willighagen. Chemical Markup, XML, and the World Wide Web. 5. Applications of Chemical Metadata in RSS Aggregators. *J. Chem. Inf. Comput. Sci.*, 44:462–469, 2004.