Chemistry Central Journal

Poster presentation

Open Access

Parsers for SMILES and SMARTS

Andrew Dalke

Address: Dalke Scientific Software, Rosensköldsgatan 9, Göteborg 412 58, Sweden

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

Published: 26 March 2008

Chemistry Central Journal 2008, **2**(Suppl 1):P40 doi:10.1186/1752-153X-2-S1-P40

This abstract is available from: http://www.journal.chemistrycentral.com/content/2/S1/P40

SMILES [1] and SMARTS [2] are two line notations developed by Daylight and implemented in a number of chemical informatics software tools. For the most part these are hand-written parsers and quite complicated and hard to read, modify, maintain, optimize or reuse. A traditional computer science approach would use a parsing system like lex/yacc but that has not made much inroad in computational chemistry. The tools have been difficult to use, especially for error reporting and recovery, and most of the developers have a chemistry background and don't know the language theory underlying this approach.

Modern programming languages and parser systems have made many of the difficulties disappear. I have been working with people from OpenSMILES [3] and several of the existing open source toolkits (Open Babel [4], CDK [5] and RDKit [6]) develop valid, useful grammars for SMILES and SMARTS. I have also been evaluating how to implement those grammars using parsing systems like ANTLR [7], PLY [8] and ragel [9]. My plan is to fold that work back into the different projects so there is a broader and more consistent support for these two important notations. I expect also that resulting code will be faster, more maintainable, and more flexible for trying new ideas. By documenting the different parts I hope the knowledge of how to use parser frameworks is disseminated into the computational chemistry development community and helps to develop the next generation of chemistry toolkits and line notations like MQL[10].

This poster presents some of the preliminary results of that work including a SMILES grammar, implementations for ANTLR and PLY, and early performance analysis.

References

1. Weininger D: *J Chem Inf Comput Sci* 1988, **28:**31-36.
2. [http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html].
3. [http://opensmiles.org/].
4. [http://openbabel.sourceforge.net/].
5. [http://almost.cubic.uni-koeln.de/cdk/cdk_top].
6. [http://www.rdkit.org/].
7. [http://antlr.org/].
8. [http://www.dabeaz.com/ply/].
9. [http://www.cs.queensu.ca/~thurston/ragel/].
10. Proschak E, Wegner J, Schüller A, Schneider G, Fechner UJ: *Chem Inf Model* 2007, **47(12):**295-301.