Journal of **Cheminformatics**

POSTER PRESENTATION

Open Access

OrChem: an open source chemistry search engine for Oracle

Mark L Rijnbeek[*], Christoph Steinbeck

From 5th German Conference on Cheminformatics: 23. CIC-Workshop
Goslar, Germany. 8-10 November 2009

Registration, indexing and searching of chemical structures in relational databases is one of the core areas of cheminformatics. However, little detail has been published on the inner workings of search engines and their development is mostly closed-source. We decided to implement an open source chemical library for Oracle, the de-facto database standard in the commercial world.

We present OrChem, an extension for the Oracle 11G database that adds registration and indexing of chemical structures to support fast substructure and similarity searching. The cheminformatics functionality is provided by the Chemistry Development Kit [1,2]. OrChem enables similarity searching with response times in the order of seconds for databases with millions of compounds, depending on provided similarity cut-off. For substructure searching, OrChem can make use of multiple processor cores on today's powerful database servers to provide fast response times in equally large data sets.

OrChem is a mix of PL/SQL and Java that executes inside the database. The user interacts with OrChem with calls to PL/SQL and Java Stored Procedures. Starting with Oracle 11g there is a just-in-time (JIT) compiler for the Oracle JVM environment which makes Java run much faster inside the database than previously.

OrChem is built on top of the Chemistry Development Kit (CDK) and depends on this Java library in numerous ways. For example, compounds are represented internally as CDK molecule objects, the CDK's I/O package is used to retrieve compound data, and its subgraph isomorphism algorithms are used for substructure validation. OrChem adds its own Java layer on top of the CDK to implement fast database storage and retrieval. With the CDK loaded into Oracle, a large cheminformatics library becomes readily available to PL/SQL. With little effort developers can build database

functions around the CDK and so quickly implement chemistry extensions for Oracle.

Published: 4 May 2010

References
1. Steinbeck, *et al*: *J Chem Inf Comput Sci* 2003, **43**(2):493-500.
2. Steinbeck, *et al*: *Curr Pharm Des* 2006, **12**(17):2111-2120.

EMBL EBI, Wellcome Trust Genome Campus, Cambridge, UK