chem-bla-ics

Cheminformatics Benchmark Project #1

Egon Willighagen

Published December 5, 2008

Citation

Willighagen, E. (2008). Cheminformatics Benchmark Project #1. In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/wngzb-y3329

Keywords

Cheminf, Cdk

Yesterday's blog about Who says Java is not fast?!? caused quite some feedback (thanx to all commenters!) with several good points. Of course, a table like that in the cinfony paper (see also the comments in the blogs by Noel (the author) and Rich). Many things determine why the CDK might be fastest in that table for SDF iterating. Suggestions have been that OpenBabel and RDKit may be doing much more than simple reading; Java might actually take advantage of the second core for caching file content.

ZZ observed something I overlooked: calculating the molecular mass in CDK is by far slowest of all three toolkit, though people have suggestions on why that may is.

Benchmarking

The correct way to compare toolkits, open source, proprietary, free, commercial, is to have a proper benchmark toolkit for cheminformatics. That's what I am suggesting here: a project to define simple and fair benchmarks. It's an open project, and anyone can contribute in order to keep tests balanced in impartial towards any tested toolkit.