

POSTER PRESENTATION

Open Access

The FPS fingerprint format and chemfp toolkit

Andrew Dalke

From 8th German Conference on Chemoinformatics: 26 CIC-Workshop
Goslar, Germany. 11-13 November 2012

During GCC 2010 poster session I presented a draft version of the FPS format for storing dense binary fingerprints. That format is now stable, and supported by RDKit [1], CACTVS [2], and other software. The chemfp package is a set of command-line tools and a Python library for fingerprint generation and high-speed Tanimoto search. It can extract pre-computed fingerprints from an SD tag or use OpenEye's OEChem [3], Open Babel [4], or RDKit to generate fingerprints. Search uses a combination of careful indexing [5], CPU-specific instructions (if available), and OpenMP. Nearest-100 similarity searches of PubChem-sized take less than a second on a laptop, and Butina clustering [6] of 2 million compounds takes about 6 hours on a 15 CPU node. In my poster I present the FPS format and chemfp package, and describe how the memory and performance requirements lead to the internal search architecture.

Published: 22 March 2013

References

1. [http://rdkit.org].
2. [http://xemistry.org/].
3. [http://www.eyesopen.com/oechem-tk].
4. [http://openbabel.org].
5. Swamidass SJ, Baldi P: **Bounds and Algorithms for Fast Exact Searches of Chemical Fingerprints in Linear and Sublinear Time.** *J Chem Inf Model* 2007, **47**:302-317.
6. Butina D: **Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets.** *J Chem Inf Model* 1999, **39**:747-750.

doi:10.1186/1758-2946-5-S1-P36

Cite this article as: Dalke: The FPS fingerprint format and chemfp toolkit. *Journal of Cheminformatics* 2013 **5**(Suppl 1):P36.

Correspondence: dalke@dalkescientific.com
Andrew Dalke Scientific, 41134, Göteborg, Sweden

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
<http://www.chemistrycentral.com/manuscript/>


ChemistryCentral