# One central tool for Chemoinformatic

RDKIT UGM 2016

DR. GUILLAUME GODIN

PRINCIPAL SCIENTIST FIRMENICH

# Firmenich an family 120 year old company


10 year ago, I started as PostDoc

Introduce RDKit 7 years ago in my group

# One central tool to do what ?

- Read molecules (sdf, smi,**canonical smiles**,…)
- Flexible descriptors
- Optimized 3D & SCF
  - MMFF94 (rdkit, conformers)
  - AM1 (mopac, molds)
  - PM3/6 (mopac)
  - Get Mulliken charges
- Get 3D descriptors:
  - Dragon like
- Get Fingerprints:
  - Neural networks FP
- Save Data:
  - MySQL, ORACLE, files, Matlab,…

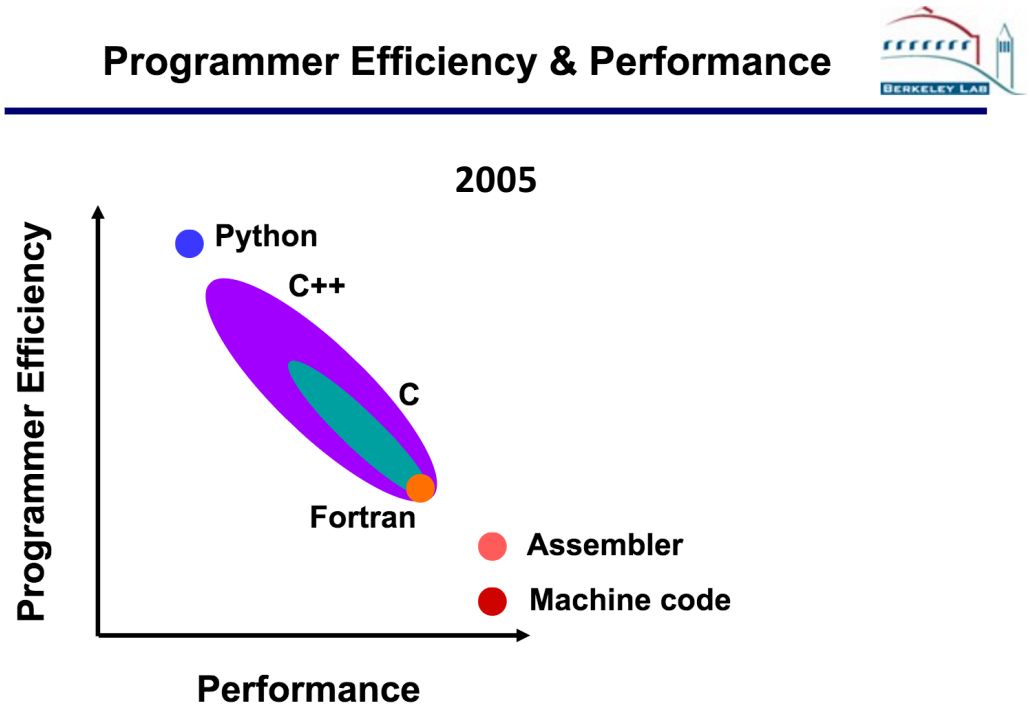
# Why RDKit ?

- Very flexible
- Very supported
- Very Stable
- Very cheap ;-)
- BONUS: Already large toolkit functions
- And...

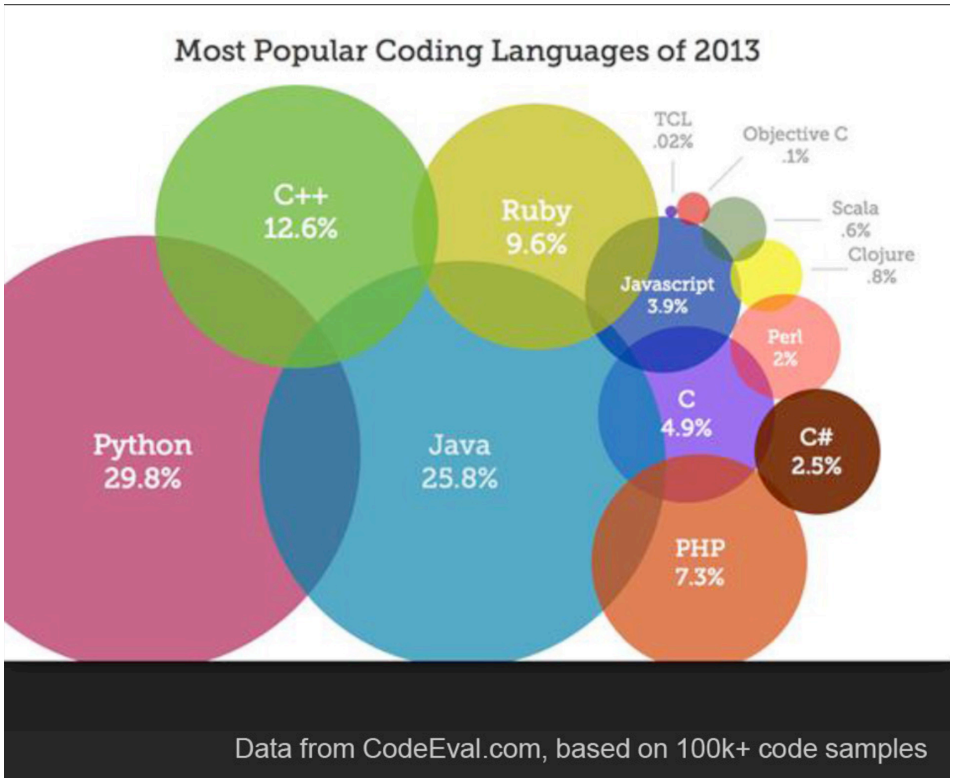
# In progress on my fork

- 3D Descriptors port from Dragon to RDKit:
  - 3D autocorrelations (done)
  - GETAWAY (done)
  - RDF (done)
  - Morse (done)
  - WHIM (in progress)

# Language vs speed coding vs speed processing



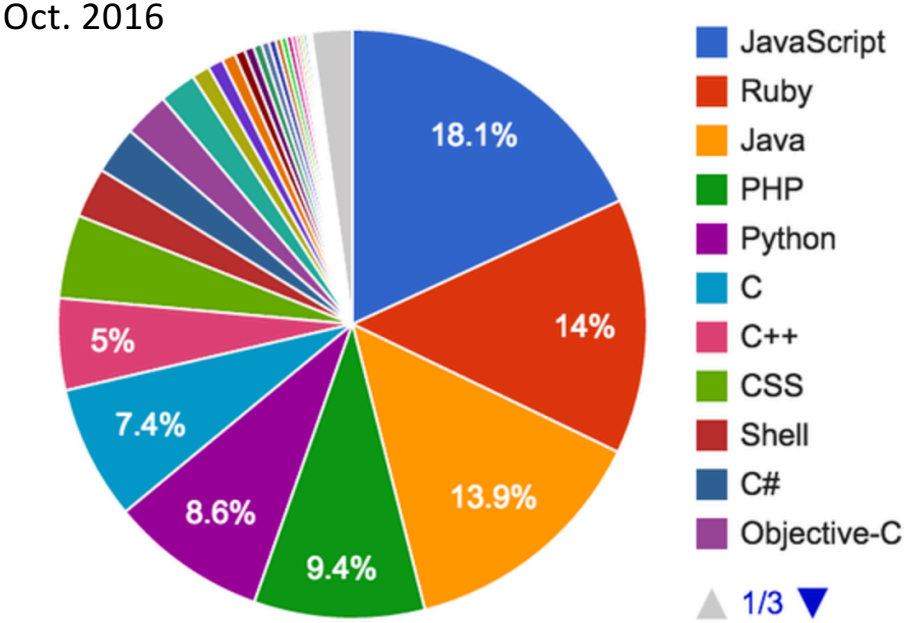

[Source: Spectrum of languages Ralf W. Grosse-Kunstleve Computational ...]



# Future…

According to GitHub, every other language has lost to JavaScript and not just Python.

Oct. 2016

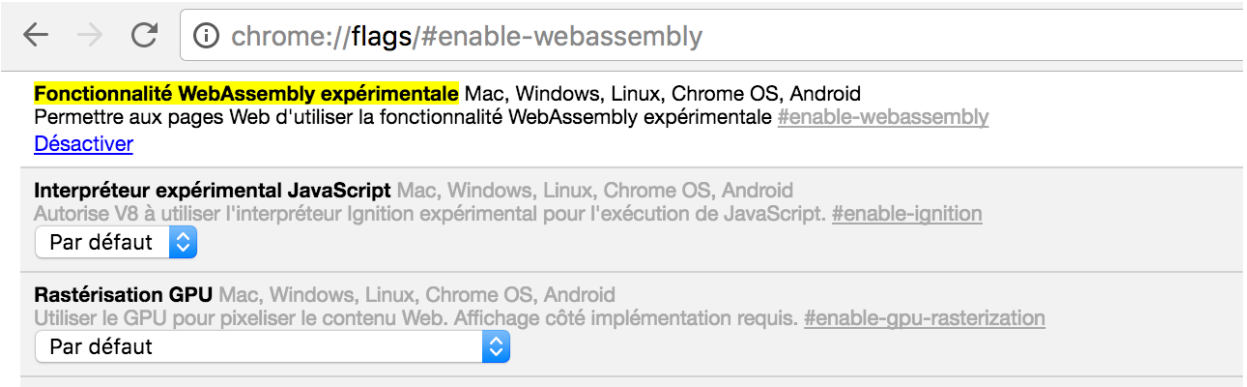


# RDKit.js implementation

Code hosted on friend group at EPFL:

https://github.com/cheminfo/RDKitjs

- In Hibernation:
  - Waiting WebAssembly for performance
  - http://webassembly.org/stepbystep/
  - It's coming … but



# … A more complex challenge

Fast semi empirical for Charges Mulliken distribution and/or 3D geometry conformers

- Initial works on molds opensource package …
  - Implementing new Atoms not on the original code (P,S,Br,I)
  - Comparing Mopac vs Molds with new Atoms vs RDKit MMFF94 (not stable and slow)
- Go back to basic Indo / cndo / mndo fortran:
  - Open source Fortran available but from divers sources
  - Wrapping fortran in C++ to avoid passing parameters and x,y,z in files

# Thanks you!

# Thanks you!

**Greg, Bryan, Nadine, …. Paolo**

**An all others great contributors.**