RDKit UGM 2019

Hamburg, Germany

https://github.com/rdkit/UGM_2019

#RDKitUGM2019

Gold Sponsors

Open for Innovation KNIME®

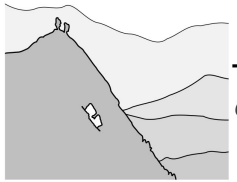
cresset® software

SCHRÖDINGER®

Silver Sponsors

RELAY THERAPEUTICS

Roche



T5 Informatics
Chemical Data Science

Bronze Sponsors

BenevolentAI

Hosting

UHH

Thanks!

Logistics

Dinners

Post-dinner socializing

Thursday morning (?) jogging

Friday morning (?) jogging

Logistics

Lightning talks: if you'd like to do one please contact Emanuel or I.

Hackathon ideas: https://github.com/rdkit/rdkit/issues labelled "Hackathon Idea"

TeachOpenCADD hackathon ideas: http://bit.ly/2kJxerQ

Job board: pinboard available, also an online version: https://bit.ly/2m4nqZX

Current lightning talk proposals

Who	What	When
Ruud van Deursen	GEN: Highly Efficient SMILES Explorer Using Autodidactic Generative Examination Networks https://arxiv.org/abs/1909.04825	Wed
Franziska Kruger	Chemical series: classifying like a medicinal chemist	Wed
Andrew Dalke	mmpdb crowdfunding	Wed
Kathryn Loving	RDKit microservices (why & how)	Wed
Ian Wetherbee	RDKit + Google Patents	Wed
Dan Nealschneider	Schrödinger contributions to the RDKit build system. We've got your back!	Wed
Mike Beachy	Ketcher + RDKit (Quick Demo)	Thurs
Guillaume Godin	Reproducible AI!?	Thurs
Matthew Sellwood	MMP graph database	Thurs
Paolo Tosco	An interesting MCS case	Thurs
Andrew Dalke	Ideas	Thurs
Greg Landrum	Something new	Thurs

Hackathon (Friday)

- You don’t need to be an expert
- Come and work on something interesting/useful.
  - Code
  - Documentation
  - Examples
  - KNIME workflows
  - TeachOpenCADD lessons
- Typically small groups working together
- Not particularly structured (i.e. groups need to self-organize)
- Not comfortable working on something on your own? Find someone else to work with/learn from

Questions

- Who is planning on attending the hackathon on Friday?

Introductions

Who am I?

Greg Landrum (T5 Informatics GmbH, KNIME AG)

Who organized this?

Who are you?

Introductions

Who am I?

Who organized this?

Emanuel Ehmki, Anna Lina Heinzke, Theresa Cavasin, Uschi Dolfus, Tim Kuhrt (U Hamburg)

Who are you?

Special thanks to Prof. Matthias Rarey and the Uni Hamburg

Aside: Past meetings/organizers

1. London (ICR): Nathan Brown
2. Hinxton (EMBL): George Papadatos
3. Darmstadt (Merck): Paul Czodrowski
4. Zurich (ETH): Sereina Riniker
5. Basel (Novartis/Roche): Nadine Schneider, Nik Stiefl, Christian Kramer
6. Berlin (Charité Berlin/FU Berlin): Andrea Volkamer, Gerhard Wolber
7. Cambridge (Cambridge University): Andreas Bender

Introductions

Who am I?

Who organized this?

Who are you?

- Industry, Academic, Other
  - Industry: Pharma/Biotech, Software, Other
  - Academic: Student, Postdoc, Faculty, Other
- Europe, North America, Asia, Other

Introductions

Who am I?

Who organized this?

Who are you?

- First RDKit UGM?
- 3 or more RDKit UGMs?
- 5 or more RDKit UGMs?

“Frequent Hitters”

- Eight UGMs:
  - Roger S
  - Gregori G
- Seven UGMs:
  - Sereina R
  - Paul C

Introductions

Who am I?

Who organized this?

Who are you? How are you using the RDKit?

- From Python
- From C++
- From Java
- From C#
- From KNIME
- From PostgreSQL
- Other
- Not sure yet

Introductions

Who am I?

Who organized this?

Who are you? Contributors

- Code
- Documentation
- Bug reports
- On the mailing list
- Other

About the meeting

- Friendly
- Diverse
  - People
  - Topics
  - Degrees of technicality
- By/for the community
  - Organized and run by volunteers

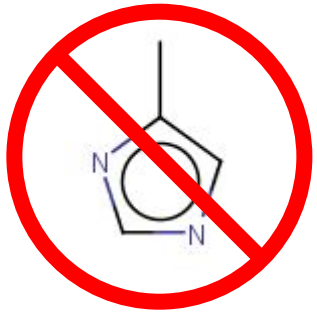
A request: social networking and this meeting

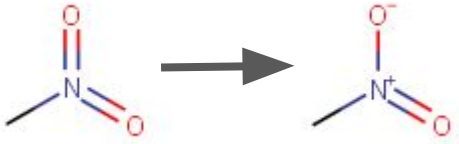
- Speakers/poster presenters: if you don’t want people posting about your contribution and/or posting pictures, please just say something (or include a note on your first slide/post-it note near your poster, or just ask the moderator for your session)
- Audience: please honor these requests

A quick aside about the RDKit

Being opinionated

- The RDKit is not designed to be a toolkit for file format conversion, so round-tripping molecules isn't always possible
- The default settings will reject molecules that are chemically unreasonable. You can turn this off[1]
- The toolkit generally does not try to guess and "fix" input structures





[1] But you generally shouldn't!

Next topic: community

Community

The heart of any *successful* open-source project

Sustainability: the bus problem

https://commons.wikimedia.org/wiki/File:Postauto_susten.jpg

Support

- Web searches
- Mailing list
- Github

- Commercial support

Community

- Mailing lists: >850 messages to rdkit-discuss from 2018.08.15 - 2019.08.15
- Google scholar: >550 hits for "rdkit" in 2018
- Searching github for `from rdkit import Chem` returns >8500 code results across >450 unique repo names (i.e. not RDKit forks).
- Each of the last five UGMs at capacity with 40-100+ attendees
- Four Google Summer of Code projects over the last three years

Code contributions in the last year

Contributions to github issue tracker in the last year

ricrogz bp-kelley ombanck UnixJunkie shayakhmetov LivC182 kovasap kienerj coleb yurivict tdudgeon tawe141 sroughley soerendip sbhakat rmrmg paconius gedeck ericmjl e-kwsm b-mahjour aparente-nurix SiPa13 CamAnNguyen yshen22 yphillip yamasakih xiaohongniua wtriddle uditgupta0912 timholy thegodone tduigou stefdoerr smoe simonmb sihagmnis36 shashany sdvillal sahertariq07 rvianello pstjohn pschwllr proteneer poppy7675 poganyp pavlovnicola oivulf mwojcikowski msteijaert mjw99 mayankBIL malteseunderdog lorton lilleswing likhangy lewisacidic kexul kennethriva kemaeleon karolbadowski jwarmitage jones-gareth jasad1 icamps hsiaoyi0504 hjuinj grinnnnnn goraj gncs ghiandonigianmarco gauravmoghe felixekn eugene-bright ericmjonas eloyfelix dvidmon dpwildboar darkcircle danpol cwhidden cowsandmilk complext clarezhu cing chazanov btcooooper bembel balducci baerbock azedine-healx andt88 andrewtarzia agdecm adalke ZacharyKaplan ValeryPolyakov Szirenke SamuelFigueroa SRaent Plancalkuele NadineSchneider Mickdub Mario-Liu JLVarjo Dekken ChinzoD ChiCheng45 CKannas Bjoux2 BillLawrence111 AustinApple Andy-Wilkinson 7FeiW

That's 115 different people

Usage in other open-source projects

- stk (docs, paper) - a Python library for building, manipulating, analyzing and automatic design of molecules.
- OpenFF - Open source approach for better force fields
- gpusimilarity - GPU implementation of fingerprint similarity searching
- Samson Connect - Software for adaptive modeling and simulation of nanosystems
- mol_frame - Chemical Structure Handling for Dask and Pandas DataFrames
- mmpdb 2.0 - matched molecular pair database generation and analysis
- CheTo - Chemical topic modeling
- OCEAN - web-tool for target-prediction of chemical structures which uses ChEMBL as datasource
- Coot - software for macromolecular model building, model completion and validation
- DeepChem - deep learning toolkit for drug discovery
- sdf_viewer.py - an interactive SDF viewer
- sdf2ppt - Reads an SDFile and displays molecules as image grid in powerpoint/openoffice presentation.
- chemfp
- PYPL - Simple cartridge that lets you call Python scripts from Oracle PL/SQL.
- WONKA - Tool for analysis and interrogation of protein-ligand crystal structures
- OOMMPPAA - Tool for directed synthesis and data analysis based on protein-ligand crystal structures
- RRDKit - RDKit integration for R
- chemicalite - SQLite integration for the RDKit
- django-rdkit - Django integration for the RDKit
- … more ...

Usage in commercial tools

- Cresset Software
- Dalke Scientific Software
- NextMove Software
- Schrödinger
- SCM
- Wolfram Research

Disclaimer: this info is from public statements made by people from those companies.
I almost certainly have forgotten someone

Usage in online tools/resources

- ChEMBL
- ZINC
- Google Patents
- PDBe
- Enamine
- TeachOpenCADD (teaching material)

Disclaimer: this info is from public statements made by people associated with those projects. I almost certainly have forgotten someone

Ok, enough of that, let’s look at what’s new.