

Bachelor Thesis Topics 2013/1

Dr. Michael Emmerich

Genetic Portfolio Search Engine



- Portfolio subset selection is used frequently in finance to invest in assets in the stock market
- However it can be applied to a broader class of problems:
 - Drug Discovery in Databases: Which chemical substances from a catalogue should be tested, maximizing the probability of a success and minimizing the risk to not discover a drug.
 - Search Engines: Decide which web-pages should be presented to a busy user, maximizing the probability of matching user preferences and minimizing the risk of failure.
- The portfolio subset selection problem in these cases is a mixed integer optimization problem that can be solved with quadratic programming (QP). However, for large sets QP gets prohibitively slow.
- The goal of this project is to design and implement a Genetic Algorithm and compare it to an existing QP in terms of solution quality, time consumption and scalability.
- The QP method and the objective function are available.
- Required: Knowledge in (1) genetic or evolutionary algorithms and (2) MATLAB.
- Supervisors: *Dr. Michael Emmerich, Dr. Iryna Yevyeva*
- Duration: 3 month

Chemical search results for 'BB and Research Organics [7 out of 4402(alt)]'. The table lists various chemical compounds with their LogP, SMILES, Molecular weight, and Ordering Info.

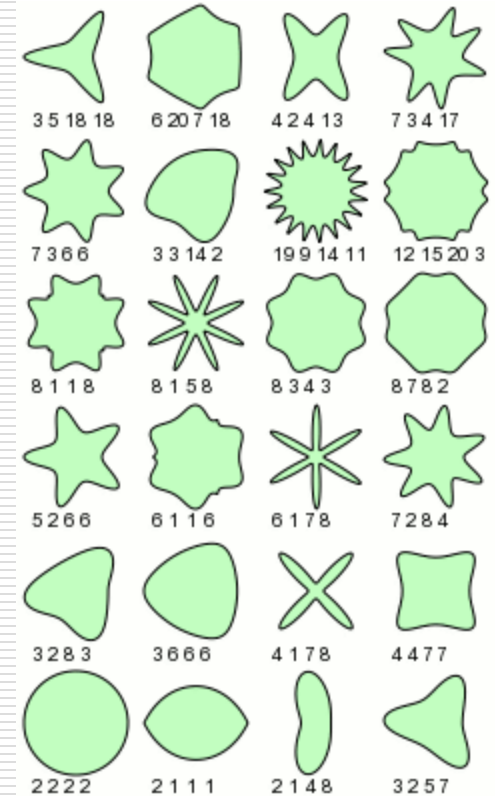
LogP #	SMILES	Molecular formula	Molecular weight	ORDERING INFO
0000001	SBB000001	CS H9 N	121.12	email: chemorder@chemdbout.com
0000002	SBB000002	CS H6 N2 O	124.14	email: chemorder@chemdbout.com
0000003	SBB000003	CS H8 N2 O	124.14	email: chemorder@chemdbout.com
0000004	SBB000004	C4 H4 F2 N2	118.09	email: chemorder@chemdbout.com
0000005	SBB000005	CS H4 N2 O2	112.09	email: chemorder@chemdbout.com
0000006	SBB000006	CS H6 N2 O2	126.11	email: chemorder@chemdbout.com
0000007	SBB000007	CS H6 N2 O2	126.11	email: chemorder@chemdbout.com
0000008	SBB000008	CS H6 N2 O2	126.11	email: chemorder@chemdbout.com
0000009	SBB000009	CS H8 N2 O2	140.14	email: chemorder@chemdbout.com
0000010	SBB000010	CS H8 N2 O2	140.14	email: chemorder@chemdbout.com
0000011	SBB000011	CS H8 F2 N2	146.14	email: chemorder@chemdbout.com

Details for the selected compound (SMILES: CC1=CN=C1C(=O)O):

- Structure: CC1=CN=C1C(=O)O
- Molecular formula: $C_5H_9N_2O_2$
- Molecular weight: 126.11
- SMILES: CC1=CN=C1C(=O)O
- ORDERING INFO: email: chemorder@chemdbout.com

ELSA on Superspheres

- ❑ Level set approximation can be used to approximate level sets
- ❑ ELSA is a simple evolutionary algorithm that does this based on a set-indicator based on K-Nearest Neighbours algorithms and/or Solow-Polasky Diversity.
- ❑ So far only 8158 superformula has been tested.
- ❑ The task is to apply ELSA on a broader test, and find strategies of improvement of this algorithm.
- ❑ Supervisors: Michael Emmerich, Andre Deutz, Iryna Yevseyeva
- ❑ We provide ELSA basic code.



kills

