NAME

RDKitGenerateMolecularFrameworks.py - Generate Bemis Murcko molecular frameworks

SYNOPSIS

RDKitGenerateMolecularFrameworks.py [--infileParams <Name,Value,...>] [--mode <GraphFrameworks or AtomicFrameworks>] [--outfileParams <Name,Value,...>] [--overwrite] [--removeDuplicates <yes or no>] [--sort <yes or no>] [--sortOrder <ascending or descending>] [--useChirality <yes or no>] [-w <dir>] -i <infile> -o <outfile>

RDKitGenerateMolecularFrameworks.py -h | --help | -e | --examples

DESCRIPTION

Generate Bemis Murcko [Ref 133] molecular frameworks for molecules. Two types of molecular frameworks can be generated: Graph or atomic frameworks. The graph molecular framework is a generic framework. The atom type, hybridization, and bond order is ignore during its generation. All atoms are set to carbon atoms and all bonds are single bonds. The atom type, hybridization, and bond order is preserved during generation of atomic molecular frameworks.

The supported input file formats are: SD (.sdf, .sd), SMILES (.smi, .csv, .tsv, .txt)

The supported output file formats are: SD (.sdf, .sd), SMILES (.smi)

OPTIONS

-e, --examples

Print examples.

-h, --help

Print this help message.

-i, --infile <infile>

Input file name.

--infileParams <Name,Value,...> [default: auto]

A comma delimited list of parameter name and value pairs for reading molecules from files. The supported parameter names for different file formats, along with their default values, are shown below:

```
SD: removeHydrogens,yes,sanitize,yes,strictParsing,yes
SMILES: smilesColumn,1,smilesNameColumn,2,smilesDelimiter,space,
    smilesTitleLine,auto,sanitize,yes
```

Possible values for smilesDelimiter: space, comma or tab.

-m, --mode <GraphFrameworks or AtomicFrameworks> [default: GraphFrameworks]

Type of molecular frameworks to generate for molecules. Possible values: GraphFrameworks or AtomicFrameworks. The graph molecular framework is a generic framework. The atom type, hybridization, and bond order is ignore during its generation. All atoms are set to carbon atoms and all bonds are single bonds. The atom type, hybridization, and bond order is preserved during the generation of atomic molecular frameworks.

-o, --outfile <outfile>

Output file name.

--outfileParams <Name,Value,...> [default: auto]

A comma delimited list of parameter name and value pairs for writing molecules to files. The supported parameter names for different file formats, along with their default values, are shown below:

```
SD: compute2DCoords,auto
SMILES: smilesDelimiter,space,smilesTitleLine,yes
```

Default value for compute2DCoords: yes for SMILES input file; no for all other file types.

--overwrite

Overwrite existing files.

-r, --removeDuplicates <yes or no> [default: no]

Remove duplicate molecular frameworks. Possible values: yes or no. The duplicate molecular franworks are identified using canonical SMILES. The removed frameworks are written to a separate output file.

-s, --sort <yes or no> [default: no]

Sort molecular frameworks by heavy atom count. Possible values: yes or no.

--sortOrder <ascending or descending> [default: ascending]

Sorting order for molecular frameworks. Possible values: ascending or descending.

-u, --useChirality <yes or no> [default: yes]

Use stereochemistry for generation of canonical SMILES strings to identify duplicate molecular frameworks.

-w, --workingdir <dir>

Location of working directory which defaults to the current directory.

EXAMPLES

To generate graph molecular framworks for molecules and write out a SMILES file, type:

```
% RDKitGenerateMolecularFrameworks.py -i Sample.smi -o SampleOut.smi
```

To generate graph molecular framworks, remove duplicate frameworks for molecules and write out SD files for unique and duplicate frameworks, type:

```
% RDKitGenerateMolecularFrameworks.py -m GraphFrameworks -r yes
  -i Sample.sdf -o SampleOut.sdf
```

To generate atomic molecular framworks, remove duplicate frameworks, sort framworks by heavy atom count in ascending order, write out SMILES files for unique and duplicate frameworks, type:

```
% RDKitGenerateMolecularFrameworks.py -m AtomicFrameworks -r yes
  -s yes -i Sample.smi -o SampleOut.smi
```

To generate graph molecular framworks for molecules in a CSV SMILES file, SMILES strings in column 1, name in olumn 2, emove duplicate frameworks, sort framworks by heavy atom count in decending order and write out a SD file, type:

```
% RDKitGenerateMolecularFrameworks.py -m AtomicFrameworks
  --removeDuplicates yes -s yes --sortOrder descending --infileParams
  "smilesDelimiter,comma,smilesTitleLine,yes,smilesColumn,1,
  smilesNameColumn,2" --outfileParams "compute2DCoords,yes"
  -i SampleSMILES.csv -o SampleOut.sdf
```

AUTHOR

Manish Sud(msud@san.rr.com)

SEE ALSO

RDKitConvertFileFormat.py, RDKitDrawMolecules.py, RDKitSearchFunctionalGroups.py, RDKitSearchSMARTS.py

COPYRIGHT

Copyright (C) 2018 Manish Sud. All rights reserved.

The functionality available in this script is implemented using RDKit, an open source toolkit for cheminformatics developed by Greg Landrum.

This file is part of MayaChemTools.

MayaChemTools is free software; you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License as published by the Free Software Foundation; either version 3 of the License, or (at your option) any later version.