NAME

RDKitRemoveSalts.py - Remove salts

SYNOPSIS

RDKitRemoveSalts.py [--infileParams <Name,Value,...>] [--mode <remove or count>] [--outfileParams <Name,Value,...>] [--overwrite] [--saltsMode <ByComponent, BySMARTSFile, BySMARTS>] [--saltsFile <FileName or auto>] [--saltsSMARTS <SMARTS>] [-w <dir>] [-o <outfile>] -i <infile>

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

DESCRIPTION

Remove salts from molecules or simply count the number of molecules containing salts. Salts are identified and removed based on either SMARTS strings or by selecting the largest disconnected components in molecules as non-salt portion of molecules.

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 <remove or count> [default: remove]

Specify whether to remove salts from molecules and write out molecules or or simply count the number of molecules containing salts.

-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,kekulize,no
SMILES: kekulize,no,smilesDelimiter,space, smilesIsomeric,yes,
    smilesTitleLine,yes
```

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

--overwrite

Overwrite existing files.

-s, --saltsMode <ByComponent, BySMARTSFile, BySMARTS> [default: ByComponent]

Specify whether to identify and remove salts based on SMARTS strings or by selecting the largest disconnected component as non-salt portion of a molecule. Possible values: ByComponent, BySMARTSFile or BySMARTS.

--saltsFile <FileName or auto> [default: auto]

Specify a file name containing specification for SMARTS corresponding to salts or use default salts file, Salts.txt, available in RDKit data directory. This option is only used during 'BySMARTSFile' value of '-s, --saltsMode' option.

RDKit data format: Smarts<tab>Name(optional)

For example:

```
[Cl,Br,I]
[N](=O)(O)O
[CH3]C(=O)O    Acetic acid
```

--saltsSMARTS <SMARTS text>

>   Space delimited SMARTS specifications to use for salts identification instead their specifications in '--saltsFile'. This option is only used during 'BySMARTS' value of '-s, --saltsMode' option.

-w, --workingdir <dir>

>   Location of working directory which defaults to the current directory.

EXAMPLES

To remove salts from molecules in a SMILES file by keeping largest disconnected components as non-salt portion of molecules and write out a SMILES file, type:

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

To count number of molecule containing salts from in a SD file, using largest components as non-salt portion of molecules, without generating any output file, type:

```
% RDKitRemoveSalts.py -m count -i Sample.sdf
```

To remove salts from molecules in a SMILES file using SMARTS strings in default Salts.txt distributed with RDKit to identify salts and write out a SMILES file, type:

```
% RDKitRemoveSalts.py -m remove -s BySMARTSFile -i Sample.smi
  -o SampleOut.smi
```

To remove salts from molecules in a SD file using SMARTS strings in a local CustomSalts.txt to identify salts and write out a SMILES file, type:

```
% RDKitRemoveSalts.py -m remove -s BySMARTSFile --saltsFile
  CustomSalts.txt -i Sample.sdf -o SampleOut.smi
```

To remove salts from molecules in a SD file using specified SMARTS to identify salts and write out a SD file, type:

```
% RDKitRemoveSalts.py -m remove -s BySMARTS  --saltsSMARTS
  '[Cl,Br,I]  [N](=O)(O)O [N](=O)(O)O'
  -i Sample.sdf -o SampleOut.smi
```

To remove salts form molecules from a CSV SMILES file, SMILES strings in column 1, name in column 2, and generate output SD file, type:

```
% RDKitRemoveSalts.py --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, RDKitRemoveDuplicateMolecules.py, RDKitSearchFunctionalGroups.py, RDKitSearchSMARTS.py

COPYRIGHT