NAME

RDKitUtil

SYNOPSIS

import RDKitUtil

DESCRIPTION

RDKitUtil module provides the following functions:

GetMolName, IsMolEmpty, MoleculesWriter, ReadAndValidateMolecules, ReadMolecules, ReadMoleculesFromMol2File, ReadMoleculesFromMolFile, ReadMoleculesFromPDBFile, ReadMoleculesFromSDFile, ReadMoleculesFromSMILESFile, WriteMolecules

FUNCTIONS

GetMolName

```
GetMolName(Mol, MolNum = None)
```

Get molecule name.

Arguments:

```
Mol (object): RDKit molecule object.
MolNum (int or None): Molecule number in input file.
```

Returns:

```
str : Molname corresponding to _Name property of a molecule, generated
    from specieid MolNum using the format "Mol%d" % MolNum, or an
    empty string.
```

IsMolEmpty

```
IsMolEmpty(Mol)
```

Check for the presence of atoms in a molecule.

Arguments:

```
Mol (object): RDKit molecule object.
```

Returns:

```
bool : True - No atoms in molecule; Otherwise, false.
```

MoleculesWriter

```
MoleculesWriter(FileName, **KeyWordArgs)
```

Set up a molecule writer.

Arguments:

```
FileName (str): Name of a file with complete path.
**KeyWordArgs (dictionary) : Parameter name and value pairs for writing and
    processing molecules.
```

Returns:

```
RDKit object : Molecule writer.
```

The file extension is used to determine type of the file and set up an appropriate file writer.

ReadAndValidateMolecules

```
ReadAndValidateMolecules(FileName, **KeyWordArgs)
```

Read molecules from an input file, validate all molecule objects, and return a list of valid and non-valid molecule objects along with their counts.

Arguments:

```
FileName (str): Name of a file with complete path.
**KeyWordArgs (dictionary) : Parameter name and value pairs for reading and
    processing molecules.
```

Returns:

```
list : List of valid RDKit molecule objects.
int : Number of total molecules in input file.
int : Number of valid molecules in input file.
```

The file extension is used to determine type of the file and set up an appropriate file reader.

ReadMolecules

```
ReadMolecules(FileName, **KeyWordArgs)
```

Read molecules from an input file without performing any validation and creation of molecule objects.

Arguments:

```
FileName (str): Name of a file with complete path.
**KeyWordArgs (dictionary) : Parameter name and value pairs for reading and
    processing molecules.
```

Returns:

```
list : List of RDKit molecule objects.
```

The file extension is used to determine type of the file and set up an appropriate file reader.

ReadMoleculesFromMol2File

```
ReadMoleculesFromMol2File(FileName, Sanitize = True, RemoveHydrogens = True)
```

Read molecule from a Tripos Mol2 file.

Arguments:

```
FileName (str): Name of a file with complete path.
Sanitize (bool): Sanitize molecules.
RemoveHydrogens (bool): Remove hydrogens from molecules.
```

Returns:

```
list : List of RDKit molecule objects.
```

ReadMoleculesFromMolFile

```
ReadMoleculesFromMolFile(FileName, Sanitize = True, RemoveHydrogens = True, StrictParsing = True)
```

Read molecule from a MDL Mol file.

Arguments:

```
FileName (str): Name of a file with complete path.
Sanitize (bool): Sanitize molecules.
RemoveHydrogens (bool): Remove hydrogens from molecules.
StrictParsing (bool): Perform strict parsing.
```

Returns:

```
list : List of RDKit molecule objects.
```

ReadMoleculesFromPDBFile

```
ReadMoleculesFromPDBFile(FileName, Sanitize = True, RemoveHydrogens = True)
```

Read molecule from a PDB file.

Arguments:

```
FileName (str): Name of a file with complete path.
Sanitize (bool): Sanitize molecules.
RemoveHydrogens (bool): Remove hydrogens from molecules.
```

Returns:

```
list : List of RDKit molecule objects.
```

ReadMoleculesFromSDFile

```
ReadMoleculesFromSDFile(FileName, Sanitize = True, RemoveHydrogens = True, StrictParsing = True)
```

Read molecules from a SD file.

Arguments:

```
FileName (str): Name of a file with complete path.
Sanitize (bool): Sanitize molecules.
RemoveHydrogens (bool): Remove hydrogens from molecules.
StrictParsing (bool): Perform strict parsing.
```

Returns:

```
list : List of RDKit molecule objects.
```

ReadMoleculesFromSMILESFile

```
ReadMoleculesFromSMILESFile(FileName, SMILESDelimiter = ' ', SMILESColIndex = 0,
SMILESNameColIndex = 1, SMILESTitleLine = 1, Sanitize = 1)
```

Read molecules from a SMILES file.

Arguments:

```
SMILESDelimiter (str): Delimiter for parsing SMILES line
SMILESColIndex (int): Column index containing SMILES string.
SMILESNameColIndex (int): Column index containing molecule name.
SMILESTitleLine (int): Flag to indicate presence of title line.
Sanitize (int): Sanitize molecules.
```

Returns:

```
list : List of RDKit molecule objects.
```

WriteMolecules

```
WriteMolecules(FileName, Mols, **KeyWordArgs)
```

Write molecules to an output file.

Arguments:

```
FileName (str): Name of a file with complete path.
Mols (list): List of RDKit molecule objects.
**KeyWordArgs (dictionary) : Parameter name and value pairs for writing and
    processing molecules.
```

Returns:

```
int : Number of total molecules.
int : Number of processed molecules written to output file.
```

The file extension is used to determine type of the file and set up an appropriate file writer.

AUTHOR

Manish Sud <msud@san.rr.com>

COPYRIGHT

The functionality available in the 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.