
NAME

SplitSDFfiles.pl - Split SDFfile(s) into multiple SD files

SYNOPSIS

SplitSDFfiles.pl SDFfile(s)...

SplitSDFfiles.pl [-c, --CmpdsMode DataField | MolName | RootPrefix] [-d, --DataField DataFieldName] [-h, --help] [-m, --mode Cmpds | Files] [-n, --numfiles number] [--numcmpds number] [-o, --overwrite] [-r, --root rootname] [-w, --workingdir dirname] SDFfile(s)...

DESCRIPTION

Split *SDFfile(s)* into multiple SD files. Each new SDFfile contains a compound subset of similar size from the initial file. Multiple *SDFfile(s)* names are separated by space. The valid file extensions are *.sdf* and *.sd*. All other file names are ignored. All the SD files in a current directory can be specified either by **.sdf* or the current directory name.

OPTIONS

-c, --CmpdsMode *DataField | MolName | RootPrefix*

This option is only used during *Cmpds* value of *<-m, --mode>* option with specified *--numcmpds* value of 1.

Specify how to generate new file names during *Cmpds* value of *<-m, --mode>* option: use *SDFfile(s)* datafield value or molname line for a specific compound; generate a sequential ID using root prefix specified by *-r, --root* option.

Possible values: *DataField | MolName | RootPrefix | RootPrefix*. Default: *RootPrefix*.

For empty *MolName* and *DataField* values during these specified modes, file name is automatically generated using *RootPrefix*.

For *RootPrefix* value of *-c, --CmpdsMode* option, new file names are generated using by appending compound record number to value of *-r, --root* option. For example: *RootName* Cmd<RecordNumber>.sdf.

Allowed characters in file names are: a-zA-Z0-9_. All other characters in datafield values, molname line, and root prefix are ignore during generation of file names.

-d, --DataField *DataFieldName*

This option is only used during *DataField* value of *<-c, --CmpdsMode>* option.

Specify *SDFfile(s)* datafield label name whose value is used for generation of new file for a specific compound. Default value: *None*.

-h, --help

Print this help message.

-m, --mode *Cmpds | Files*

Specify how to split *SDFfile(s)*: split into files with each file containing specified number of compounds or split into a specified number of files.

Possible values: *Cmpds | Files*. Default: *Files*.

For *Cmpds* value of *-m, --mode* option, value of *--numcmpds* option determines the number of new files. And value of *-n, --numfiles* option is used to figure out the number of new files for *Files* value of *-m, --mode* option.

-n, --numfiles *number*

Number of new files to generate for each *SDFfile(s)*. Default: 2.

This value is only used during *Files* value of *-m, --mode* option.

--numcmpds *number*

Number of compounds in each new file corresponding to each *SDFfile(s)*. Default: 1.

This value is only used during *Cmpds* value of *-m, --mode* option.

-o, --overwrite

Overwrite existing files.

-r, --root *rootname*

New SD file names are generated using the root: <Root>Part<Count>.sdf. Default new file names: <InitialSDFfileName> Part<Count>.sdf. This option is ignored for multiple input files.

`-w,--workingdir dirname`

Location of working directory. Default: current directory.

EXAMPLES

To split each SD file into 5 new SD files, type:

```
% SplitSDFfiles.pl -n 5 -o Sample1.sdf Sample2.sdf
% SplitSDFfiles.pl -n 5 -o *.sdf
```

To split Sample1.sdf into 10 new NewSample*.sdf files, type:

```
% SplitSDFfiles.pl -m Files -n 10 -r NewSample -o Sample1.sdf
```

To split Sample1.sdf into new NewSample*.sdf files containing maximum of 5 compounds in each file, type:

```
% SplitSDFfiles.pl -m Cmpds --numcmpds 5 -r NewSample -o Sample1.sdf
```

To split Sample1.sdf into new SD files containing one compound each with new file names corresponding to molname line, type:

```
% SplitSDFfiles.pl -m Cmpds --numcmpds 1 -c MolName -o Sample1.sdf
```

To split Sample1.sdf into new SD files containing one compound each with new file names corresponding to value of datafield MolID, type:

```
% SplitSDFfiles.pl -m Cmpds --numcmpds 1 -c DataField -d MolID
-o Sample1.sdf
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

InfoSDFfiles.pl, JoinSDFfiles.pl, MolFilesToSD.pl, SDToMolFiles.pl

COPYRIGHT

Copyright (C) 2004-2010 Manish Sud. All rights reserved.

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.