NAME

MergeTextFilesWithSD.pl - Merge CSV or TSV TextFile(s) into SDFile

SYNOPSIS

MergeTextFilesWithSD.pl SDFile TextFile(s)…

MergeTextFilesWithSD.pl [-h, --help] [--indelim comma | semicolon] [-c, --columns colnum,…;… | collabel,…;…] [-k, --keys colkeynum;… | colkeylabel;…] [-m, --mode colnum | collabel] [-o, --overwrite] [-r, --root rootname] [-s, --sdkey sdfieldname] [-w, --workingdir dirname] SDFile TextFile(s)…

DESCRIPTION

Merge multiple CSV or TSV *TextFile(s)* into *SDFile*. Unless -k --keys option is used, data rows from all *TextFile(s)* are added to *SDFile* in a sequential order, and the number of compounds in *SDFile* is used to determine how many rows of data are added from *TextFile(s)*.

Multiple *TextFile(s)* names are separated by spaces. The valid file extensions are *.csv* and *.tsv* for comma/semicolon and tab delimited text files respectively. All other file names are ignored. All the text files in a current directory can be specified by **.csv*, **.tsv*, or the current directory name. The --indelim option determines the format of *TextFile(s)*. Any file which doesn't correspond to the format indicated by --indelim option is ignored.

OPTIONS

-h, --help

Print this help message.

--indelim *comma | semicolon*

Input delimiter for CSV *TextFile(s)*. Possible values: *comma or semicolon*. Default value: *comma*. For TSV files, this option is ignored and *tab* is used as a delimiter.

-c, --columns *colnum,…;… | collabel,…;…*

This value is mode specific. It is a list of columns to merge into *SDFile* specified by column numbers or labels for each text file delimited by ";". All *TextFile(s)* are merged into *SDFile*.

Default value: *all;all;…*. By default, all columns from TextFile(s) are merged into *SDFile*.

For *colnum* mode, input value format is: *colnum,…;colnum,…;…*. Example:

    "1,2;1,3,4;7,8,9"

For *collabel* mode, input value format is: *collabel,…;collabel,…;…*. Example:

    "MW,SumNO;SumNHOH,ClogP,PSA;MolName,Mol_Id,Extreg"

-k, --keys *colkeynum;… | colkeylabel;…*

This value is mode specific. It specifies column keys to use for merging *TextFile(s)* into *SDFile*. The column keys, delimited by ";", are specified by column numbers or labels for *TextFile(s)*.

By default, data rows from *TextFile(s)* are merged into *SDFile* in the order they appear.

For *colnum* mode, input value format is: *colkeynum, colkeynum;…*. Example:

    "1;3;7"

For *collabel* mode, input value format is: *colkeylabel, colkeylabel;…*. Example:

    "Mol_Id;Mol_Id;Cmpd_Id"

-m, --mode *colnum | collabel*

Specify how to merge *TextFile(s)* into *SDFile*: using column numbers or column labels. Possible values: *colnum or collabel*. Default value: *colnum*.

-o, --overwrite

Overwrite existing files.

-r, --root *rootname*

New SD file name is generated using the root: <Root>.sdf. Default file name: <InitialSDFileName>MergedWith<FirstTextFileName>1To<Count>.sdf.

-s, --sdkey *sdfieldname*

SDFile data field name used as a key to merge data from TextFile(s). By default, data rows from *TextFile(s)* are merged into *SDFile* in the order they appear.

-w, --workingdir *dirname*

Location of working directory. Default: current directory.

EXAMPLES

To merge Sample1.csv and Sample2.csv into Sample.sdf and generate NewSample.sdf, type:

```
% MergeTextFileswithSD.pl -r NewSample -o Sample.sdf
  Sample1.csv Sample2.csv
```

To merge all Sample*.tsv into Sample.sdf and generate NewSample.sdf file, type:

```
% MergeTextFilesWithSD.pl -r NewSample -o Sample.sdf
  Sample*.tsv
```

To merge column numbers "1,2" and "3,4,5" from Sample2.csv and Sample3.csv into Sample.sdf and to generate NewSample.sdf, type:

```
% MergeTextFilesWithSD.pl -r NewSample -m colnum -c "1,2;3,4,5"
  -o Sample.sdf Sample1.csv Sample2.csv
```

To merge column "Mol_ID,Formula,MolWeight" and "Mol_ID,ChemBankID,NAME" from Sample1.csv and Sample2.csv into Sample.sdf using "Mol_ID" as SD and column keys to generate NewSample.sdf, type:

```
% MergeTextFilesWithSD.pl -r NewSample -s Mol_ID -k "Mol_ID;Mol_ID"
  -m collabel -c "Mol_ID,Formula,MolWeight;Mol_ID,ChemBankID,NAME"
  -o Sample1.sdf Sample1.csv Sample2.csv
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

ExtractFromSDFiles.pl, FilterSDFiles.pl, InfoSDFiles.pl, JoinSDFiles.pl, JoinTextFiles.pl, MergeTextFiles.pl, ModifyTextFilesFormat.pl, SplitSDFiles.pl, SplitTextFiles.pl

COPYRIGHT