

NAME

MergeTextFiles.pl - Merge multiple CSV or TSV text files into a single text file

SYNOPSIS

MergeTextFiles.pl TextFiles...

MergeTextFiles.pl [-h, --help] [--indelim comma | semicolon] [-c, --columns colnum,...;... | collabel,...;...] [-k, --keys colnum,...;... | collabel,...;...] [-m, --mode colnum | collabel] [-o, --overwrite] [--outdelim comma | tab | semicolon] [-q, --quote yes | no] [-r, --root rootname] [-s, --startcol colnum | collabel] [--startcolmode before | after] [-w, --workingdir dirname] TextFiles...

DESCRIPTION

Merge multiple CSV or TSV *TextFiles* into first *TextFile* to generate a single text file. Unless `-k --keys` option is used, data rows from other *TextFiles* are added to first *TextFile* in a sequential order, and the number of rows in first *TextFile* is used to determine how many rows of data are added from other *TextFiles*.

Multiple *TextFiles* names are separated by space. 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 *TextFiles*. 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 first text file specified by column numbers or labels for each text file delimited by `;`. All specified text files are merged into first text file.

Default value: *all;all;...* By default, all columns from specified text files are merged into first text file.

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 colnum,...;... | collabel,...;...`

This value is mode specific. It specifies column keys to use for merging all specified text files into first text file. The column keys are specified by column numbers or labels for each text file delimited by `;`.

By default, data rows from text files are merged into first file 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 text files: using column numbers or column labels. Possible values: *colnum* or *collabel*. Default value: *colnum*.

`-o, --overwrite`

Overwrite existing files.

`--outdelim comma | tab | semicolon`

Output text file delimiter. Possible values: *comma*, *tab*, or *semicolon* Default value: *comma*.

`-q, --quote yes | no`

Put quotes around column values in output text file. Possible values: *yes* or *no*. Default value: *yes*.

`-r, --root rootname`

New text file name is generated using the root: `<Root>.<Ext>`. Default file name:

`<FirstTextFileName>1To<Count>Merged.<Ext>`. The *csv*, and *tsv* `<Ext>` values are used for comma/semicolon, and tab

delimited text files respectively.

`-s, --startcol colnum | collabel`

This value is mode specific. It specifies the column in first text file which is used for start merging other text files. For *colnum* mode, specify column number and for *collabel* mode, specify column label.

Default value: *last*. Start merge after the last column.

`--startcolmode before | after`

Start the merge before or after the `-s, --startcol` value. Possible values: *before* or *after* Default value: *after*.

`-w, --workingdir dirname`

Location of working directory. Default: current directory.

EXAMPLES

To merge Sample2.csv and Sample3.csv into Sample1.csv and generate NewSample.csv, type:

```
% MergeTextFiles.pl -r NewSample -o Sample1.csv Sample2.csv
Sample3.csv
```

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

```
% MergeTextFiles.pl -r NewSample --indelim comma --outdelim tab -o
Sample*.csv
```

To merge column numbers "1,2" and "3,4,5" from Sample2.csv and Sample3.csv into Sample1.csv starting before column number 3 in Sample1.csv and to generate NewSample.csv without quoting column data, type:

```
% MergeTextFiles.pl -s 3 --startcolmode before -r NewSample -q no
-m colnum -c "all;1,2;3,4,5" -o Sample1.csv Sample2.csv
Sample3.csv
```

To merge column "Mol_ID,Formula,MolWeight" and "Mol_ID,NAME,ChemBankID" from Sample2.csv and Sample3.csv into Sample1.csv using "Mol_ID" as a column keys starting after the last column and to generate NewSample.tsv, type:

```
% MergeTextFiles.pl -r NewSample --outdelim tab -k "Mol_ID;Mol_ID;
Mol_ID" -m collabel -c "all;Mol_ID,Formula,MolWeight;Mol_ID,NAME,
ChemBankID" -o Sample1.csv Sample2.csv Sample3.csv
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

JoinTextFiles.pl, MergeTextFilesWithSD.pl, ModifyTextFilesFormat.pl, SplitTextFiles.pl

COPYRIGHT

Copyright (C) 2004-2012 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.