NAME

SortTextFiles.pl - Sort TextFile(s) using values for a column

SYNOPSIS

SortTextFiles.pl TextFile(s)...

SortTextFiles.pl [-d, --detail infolevel] [-h, --help] [--indelim comma | semicolon] [-k, --key colnum | collabel] [--keydata numeric | alphanumeric] [-m, --mode colnum | collabel] [-o, --overwrite] [--outdelim comma | tab | semicolon] [-q, --quote yes | no] [-r, --root rootname] [-s, --sort ascending | descending] [-w, --workingdir dirname] TextFile(s)...

DESCRIPTION

Sort *TextFile(s)* using values for a key column specified by a column number or label. Only one column key can be specified for sorting. In an event of conflict during sorting process, two similar values for a column key are simply transferred to output files in order of their presence in input files. Additionally, rows with empty or inappropriate values for column key are simply placed at the end. The file 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 *TextFile(s)*. Any file which doesn't correspond to the format indicated by --indelim option is ignored.

OPTIONS

-d, --detail *infolevel*

Level of information to print about lines being ignored. Default: *1*. Possible values: *1, 2 or 3*.

-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.

-k, --key *col number | col name*

This value is mode specific. It specifies which column to use for sorting *TextFile(s)*. Possible values: *col number or col label*. Default value: *first column*.

--keydata *numeric | alphanumeric*

Data type for column key. Possible values: *numeric or alphanumeric*. Default value: *numeric*. For *alphanumeric* data values, comparison is case insensitive.

-m, --mode *colnum | collabel*

Specify how to sort text files: using column number or column label. 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 new file name: <InitialTextFileName>SortedByColumn.<Ext>. The csv, and tsv <Ext> values are used for comma/semicolon, and tab delimited text files respectively. This option is ignored for multiple input files.

-s, --sort *ascending | descending*

Sorting order for column values. Possible values: *ascending or descending*. Default value: *ascending*.

-w, --workingdir *dirname*

>   Location of working directory. Default: current directory.

EXAMPLES

To perform numerical sort in ascending order using first column values and generate a new CSV text file NewSample1.csv, type:

```
% SortTextFiles.pl -o -r NewSample1 Sample1.csv
```

To perform numerical sort in descending order using MolWeight column and generate a new CSV text file NewSample1.csv, type:

```
% SortTextFiles.pl -m collabel -k MolWeight --keydata numeric
  -s descending -r NewSample1 -o Sample1.csv
```

To perform numerical sort in ascending order using column number 1 and generate a new TSV text file NewSample1.csv, type:

```
% SortTextFiles.pl -m colnum -k 1 --keydata numeric -s ascending
  -r NewSample1 --outdelim tab -o Sample1.csv
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

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

COPYRIGHT

Copyright (C) 2018 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.