

NAME

SplitTextFiles.pl - Split CSV or TSV TextFile(s) into multiple text files

SYNOPSIS

SplitTextFiles.pl TextFile(s)...

SplitTextFiles.pl [-f, --fast] [-h, --help] [--indelim comma | semicolon] [-l, --label yes | no] [-n, --numfiles number] [-o, --overwrite] [--outdelim comma | tab | semicolon] [-q, --quote yes | no] [-r, --root rootname] [-w, --workingdir dirname] TextFile(s)...

DESCRIPTION

Split CSV or TSV *TextFile(s)* into multiple text files. Each new text file contains a subset of similar number of lines from the initial file. 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

-f, --fast

In this mode, *--indelim*, *--outdelim*, and *-q --quote* options are ignored. The format of input and output file(s) are assumed to be similar. And the text lines from input *TextFile(s)* are just transferred to output file(s) without any processing.

-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.

-l, --label yes | no

First line contains column labels. Possible values: *yes* or *no*. Default value: *yes*

-n, --numfiles number

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

-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 names are generated using the root: <Root>Part<Count>.<Ext>. Default new file names: <InitialTextFileName>Part<Count>.<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.

-w, --workingdir dirname

Location of working directory. Default: current directory

EXAMPLES

To split each CSV text files into 4 different text files type:

```
% SplitTextFiles.pl -n 5 -o Sample1.csv Sample2.csv
% SplitTextFiles.pl -n 5 -o *.csv
```

To split Sample1.tsv into 10 different CSV text files, type:

```
% SplitTextFiles.pl -n 10 --outdelim comma -o Sample1.csv
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

JoinTextFiles.pl, MergeTextFiles.pl, ModifyTextFilesFormat.pl

COPYRIGHT

Copyright (C) 2004-2008 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.