

NAME

ElementalAnalysisTextFiles.pl - Perform elemental analysis using formula column in TextFile(s)

SYNOPSIS

ElementalAnalysisTextFiles.pl TextFile(s)...

ElementalAnalysisTextFiles.pl [-c, --colmode *colnum* | *collabel*] [-d, --detail *infolevel*] [-f, --fast] [-f, --formulacol *colnum* | *collabel*] [-h, --help] [--indelim *comma* | *semicolon*] [-m, --mode *All* | "ElementalAnysis, [MolecularWeight, ExactMass]"] [-o, --overwrite] [--outdelim *comma* | *tab* | *semicolon*] [-p, --precision *number*] [-q, --quote *yes* | *no*] [-r, --root *rootname*] [-s, --startcol *colnum* | *collabel*] [--startcolmode *before* | *after*] [-v, --valuecollabels [*Name*, *Label*, [*Name*, *Label*,...]] [-w, --workingdir *dirname*] TextFile(s)...

DESCRIPTION

Perform elemental analysis using molecular formula column specified by a column number or label in *TextFile(s)*.

In addition to straightforward molecular formulas - H₂O, HCl, C₃H₇O₂N - other supported variations are: Ca₃(PO₄)₂, [PCl₄]⁺, [Fe(CN)₆]⁴⁻, C₃H₄N₂O₆+₂, Na₂CO₃.10H₂O, 8H₂S.46H₂O, and so on. Charges are simply ignored. Isotope symbols in formulas specification, including D and T, are not supported.

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

-c, --colmode *colnum* | *collabel*

Specify how columns are identified in *TextFile(s)*: using column number or column label. Possible values: *colnum* or *collabel*. Default value: *colnum*.

-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.

--fast

In this mode, the formula column specified using -f, --formulacol option is assumed to contain valid molecular formula data and initial formula validation check is skipped.

-f, --formulacol *col number* | *col name*

This value is mode specific. It specifies molecular formula column to use for performing elemental analysis on *TextFile(s)*. Possible values: *col number* or *col label*. Default value: *first column containing the word formula in its column label*.

-m, --mode *All* | "ElementalAnalysis,[MolecularWeight,ExactMass]"

Specify what values to calculate using molecular formula in *TextFile(s)*: calculate all supported values or specify a comma delimited list of values. Possible values: *All* | "ElementalAnalysis, [MolecularWeight, ExactMass]". Default: *All*

--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.

-o, --overwrite

Overwrite existing files.

--outdelim *comma* | *tab* | *semicolon*

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

-p, --precision *number*

Precision of calculated values in the output file. Default: up to 2 decimal places. Valid values: positive

`-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>ElementalAnalysis.<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, --startcol colnum | collabel`

This value is mode specific. It specifies the column in text files which is used for start adding calculated column values. 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 adding calculated column values after the `-s, --startcol` value. Possible values: *before* or *after*. Default value: *after*.

`-v --valuecollabels Name,Label,[Name,Label,...]`

Specify column labels to use for calculated values. In general, it's a comma delimited list of value name and column label pairs. Supported value names: *ElementalAnalysis*, *MolecularWeight*, and *ExactMass*. Default labels: *ElementalAnalysis*, *MolecularWeight*, and *ExactMass*.

`-w, --workingdir dirname`

Location of working directory. Default: current directory.

EXAMPLES

To perform elemental analysis, calculate molecular weight and exact mass using formulas in a column with the word Formula in its column label and generate a new CSV text file NewSample1.csv, type:

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

To perform elemental analysis using formulas in column number two, use column label Analysis for calculated data, and generate a new CSV text file NewSample1.csv, type:

```
% ElementalAnalysisTextFiles.pl --m ElementalAnalysis --formulacol 2
--valuecollabels "ElementalAnalysis,Analysis" -o -r NewSample1
Sample1.csv
```

To calculate molecular weight using formula in column label Formula with four decimal precision and generate a new CSV text file NewSample1.csv, type

```
% ElementalAnalysisTextFiles.pl --m MolecularWeight --colmode collabel
--formulacol Formula --precision 4 -o -r NewSample1 Sample1.csv
```

To calculate exact mass using formula in column label Formula with four decimal precision, adding column for exact mass right after Formula column, and generate a new CSV text file NewSample1.csv, type

```
% ElementalAnalysisTextFiles.pl --m ExactMass --colmode collabel
--formulacol Formula --precision 4 --startcolmode after
--startcol Formula -o -r NewSample1 Sample1.csv
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

AnalyzeTextFilesData.pl, InfoTextFiles.pl, ExtractFromTextFiles.pl

COPYRIGHT

Copyright (C) 2024 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.