

NAME

InfoFingerprintsTextFiles.pl - List information about fingerprints data in TextFile(s)

SYNOPSIS

InfoFingerprintsTextFiles.pl TextFile(s)...

```
InfoFingerprintsTextFiles.pl [-a, --all] [--AverageBitDensity] [--BitDensity] [-c, --count] [-c, --ColMode ColNum | ColLabel]
[-d, --detail InfoLevel] [--DataCheck] [-e, --empty] [--FingerprintsCol col number | col name] [--FingerprintsFormatMode
Internal | Specify] [--FingerprintsString Hexadecimal | Binary | RawBinary] [--FingerprintsType] [--FingerprintsStringType] [
--FingerprintsSize] [-h, --help] [--InDelim comma | semicolon] [--OnBits] [-w, --WorkingDir dirname] TextFile(s)...
```

DESCRIPTION

List information about fingerprints data in *TextFile(s)*: number of rows containing fingerprints data, type and size of fingerprint, bit density and average bit density of bit-based fingerprints, and so on.

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

-a, --all

List all the available information

--AverageBitDensity

List average bit density of bit-based fingerprints data

--BitDensity

List bit density of bit-based fingerprints data in each row

-c, --count

List number of rows containing fingerprints data. This is default behavior.

-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

--DataCheck

Validate fingerprints data specified using *--FingerprintsCol* and list information about missing and invalid data

-e, --empty

List number of rows containing no fingerprints data

--FingerprintsCol col number | col name

This value is *-c, --colmode* specific. It corresponds to column in *TextFile(s)* containing fingerprints data. Possible values: *col number* or *col label*. Default value: *first column containing the word Fingerprints in its column label*.

--FingerprintsFormatMode Internal | Specify

Specify format of fingerprints data in *TextFile(s)*: use default format which MayaChemTools fingerprint generation scripts use to write out fingerprints data or explicitly specify format of fingerprints. Possible values: *Internal* | *Specify*. Default value: *Internal*.

Internal fingerprints string format consists of four parts delimited by semicolon: <Type:StringType:Size:String>. For example:

```
"PathLength:Binary:512:010011..."
"MDLKeys166FP:Binary:166:010011..."
"MDLKeys166Count:Vector:166:0 1 2..."
```

For *Specify* value of *--FingerprintsFormatMode* option, *--FingerprintsString* is used to interpret fingerprints string.

--FingerprintsString Hexadecimal | Binary | RawBinary

Format of fingerprints string during *Specify* value of *--FingerprintsFormatMode* option. Possible values: *Hexadecimal*, *Binary*, or *RawBinary*. Default value: *none*; its value must be explicitly specified.

--FingerprintsType

List types of fingerprints data

--FingerprintsStringType

List types of fingerprint strings

--FingerprintsSize

List size of fingerprints

-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.

--OnBits

List number of on bits in bit-based fingerprints data for each row.

-w, --WorkingDir *DirName*

Location of working directory. Default: current directory

EXAMPLES

To count number of lines containing fingerprints data present in a column name containing Fingerprint substring, type:

```
% InfoFingerprintsTextFiles.pl SampleFP1.csv
```

To list all available information about fingerprints data in any internal format present in a column name containing Fingerprint substring, type:

```
% InfoFingerprintsTextFiles.pl -a SampleFP1.csv
```

To list all available information about fingerprints data in any internal format present in a column named PathLengthFingerprints, type:

```
% InfoFingerprintsTextFiles.pl -a --ColMode ColLabel --FingerprintsCol  
PathLengthFingerprints SampleFP2.csv
```

To list all available information about fingerprints data in hexadecimal bit-string format present in a column named PathLengthFingerprints, type:

```
% InfoFingerprintsTextFiles.pl -a --ColMode ColLabel --FingerprintsCol  
PathLengthFingerprints --FingerprintsFormatMode Specify  
--FingerprintsString Hexadecimal SampleFP2.csv
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

InfoFingerprintsTextFiles.pl, PathLengthFingerprints.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.