

NAME

InfoFingerprintsFiles.pl - List information about fingerprints data in SD, FP and CSV/TSV text file(s)

SYNOPSIS

InfoFingerprintsFiles.pl SDFfile(s) FPFile(s) TextFile(s)...

```
InfoFingerprintsFiles.pl [-a, --all] [--AverageBitDensity] [--BitDensity] [-c, --count] [-c, --ColMode ColNum | ColLabel] [
--DataCheck] [-d, --detail InfoLevel] [-e, --empty] [--FingerprintsCol col number | col name] [--FingerprintsField FieldLabel] [
--FingerprintsType] [--FingerprintsDescription] [--FingerprintsSize] [--FingerprintsBitStringFormat] [
--FingerprintsBitOrder] [--FingerprintsVectorValuesType] [--FingerprintsVectorValuesFormat] [-h, --help] [--InDelim
comma | semicolon] [--NumOfOnBits] [--NumOfNonZeroValues] [-w, --WorkingDir dirname] SDFfile(s) FPFile(s) TextFile(s)...
```

DESCRIPTION

List information about fingerprints data in *SD*, *FP* and *CSV/TSV* text file(s): number of rows containing fingerprints data, type of fingerprints vector, description and size of fingerprints, bit density and average bit density for bit-vector fingerprints strings, and so on.

The scripts InfoFingerprintsSDFiles.pl and InfoFingerprintsTextFiles.pl have been removed from the current release of MayaChemTools and their functionality merged with this script.

The valid *SDFfile* extensions are *.sdf* and *.sd*. All SD files in a current directory can be specified either by **.sdf* or the current directory name.

The valid *FPFile* extensions are *.fpf* and *.fp*. All FP files in a current directory can be specified either by **.fpf* or the current directory name.

The valid *TextFile* extensions are *.csv* and *.tsv* for comma/semicolon and tab delimited text files respectively. All other file names are ignored. All 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.

Format of fingerprint strings data in *SDFfile(s)*, *FPFile(s)* and *TextFile(s)* is automatically detected.

Example of *FP* file containing fingerprints bit-vector string data:

```
#
# Package = MayaChemTools 7.4
# ReleaseDate = Oct 21, 2010
#
# TimeStamp = Mon Mar 7 15:14:01 2011
#
# FingerprintsStringType = FingerprintsBitVector
#
# Description = PathLengthBits:AtomicInvariantsAtomTypes:MinLength1:...
# Size = 1024
# BitStringFormat = HexadecimalString
# BitsOrder = Ascending
#
Cmpd1 9c8460989ec8a49913991a6603130b0a19e8051c89184414953800cc21510...
Cmpd2 000000249400840040100042011001001980410c000000001010088001120...
... ..
... ..
```

Example of *FP* file containing fingerprints vector string data:

```
#
# Package = MayaChemTools 7.4
# ReleaseDate = Oct 21, 2010
#
# TimeStamp = Mon Mar 7 15:14:01 2011
#
# FingerprintsStringType = FingerprintsVector
#
# Description = PathLengthBits:AtomicInvariantsAtomTypes:MinLength1:...
# VectorStringFormat = IDsAndValuesString
# VectorValuesType = NumericalValues
#
Cmpd1 338;C F N O C:C C:N C=O CC CF CN CO C:C:C C:C:N C:CC C:CF C:CN C:
N:C C:NC CC:N CC=O CCC CCN CCO CNC NC=O O=CO C:C:C:C C:C:C:N C:C:CC...;
33 1 2 5 21 2 2 12 1 3 3 20 2 10 2 2 1 2 2 2 8 2 5 1 1 1 19 2 8 2 2 2 2
6 2 2 2 2 2 2 2 3 2 2 1 4 1 5 1 1 18 6 2 2 1 2 10 2 1 2 1 2 2 2 2 ...
Cmpd2 103;C N O C=N C=O CC CN CO CC=O CCC CCN CCO CNC N=CN NC=O NCN O=C
O C CC=O CCCC CCN CCCO CCNC CNC=N CNC=O CNC CCCC=O CCCCC CCCCNC CC...;
15 4 4 1 2 13 5 2 2 15 5 3 2 2 1 1 1 2 17 7 6 5 1 1 1 2 15 8 5 7 2 2 2 2
```

```

1 2 1 1 3 15 7 6 8 3 4 4 3 2 2 1 2 3 14 2 4 7 4 4 4 4 1 1 1 2 1 1 1 ...
... ..
... ..

```

Example of *SD* file containing fingerprints bit-vector string data:

```

... ..
... ..
$$$$
... ..
... ..
... ..
41 44 0 0 0 0 0 0 0 0 0999 V2000
-3.3652 1.4499 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
... ..
2 3 1 0 0 0 0
... ..
M END
> <CmpdID>
Cmpd1

> <PathLengthFingerprints>
FingerprintsBitVector;PathLengthBits:AtomicInvariantsAtomTypes:MinLength:MaxLength8;1024;HexadecimalString;Ascending;9c8460989ec8a49913991a6603130b0a19e8051c89184414953800cc2151082844a201042800130860308e8204d402800831048940e44281c00060449a5000ac80c894114e006321264401600846c05016446208190410805000304a10205b0100e04c0038ba0fad0209c0ca8b1200012268b61c0026a
aa0660a11014a011d46

$$$$
... ..
... ..

```

Example of CSV *Text* file containing fingerprints bit-vector string data:

```

"CompoundID","PathLengthFingerprints"
"Cmpd1","FingerprintsBitVector;PathLengthBits:AtomicInvariantsAtomTypes:MinLength:MaxLength8;1024;HexadecimalString;Ascending;9c8460989ec8a49913991a6603130b0a19e8051c89184414953800cc2151082844a201042800130860308e8204d402800831048940e44281c00060449a5000ac80c894114e006321264401..."
... ..
... ..

```

The current release of MayaChemTools supports the following types of fingerprint bit-vector and vector strings:

```

FingerprintsVector;AtomNeighborhoods:AtomicInvariantsAtomTypes:MinRadius0:MaxRadius2;41;AlphaNumericalValues;ValuesString;NR0-C.X1.BO1.H3-ATC1:NR1-C.X3.BO3.H1-ATC1:NR2-C.X1.BO1.H3-ATC1:NR2-C.X3.BO4-ATC1 NR0-C.X1.BO1.H3-ATC1:NR1-C.X3.BO3.H1-ATC1:NR2-C.X1.BO1.H3-ATC1:NR2-C.X3.BO4-ATC1 NR0-C.X2.BO2.H2-ATC1:NR1-C.X2.BO2.H2-ATC1:NR1-C.X3.BO3.H1-ATC1:NR2-C.X2.BO2.H2-ATC1:NR2-N.X3.BO3-ATC1:NR2-O.X1.BO1.H1-ATC1 NR0-C.X2.B...

```

```

FingerprintsVector;AtomTypesCount:AtomicInvariantsAtomTypes:ArbitrarySize;10;NumericalValues;IDsAndValuesString;C.X1.BO1.H3 C.X2.BO2.H2 C.X2.BO3.H1 C.X3.BO3.H1 C.X3.BO4 F.X1.BO1 N.X2.BO2.H1 N.X3.BO3 O.X1.BO1.H1 O.X1.BO2;2 4 14 3 10 1 1 1 3 2

```

```

FingerprintsVector;AtomTypesCount:SLogPAtomTypes:ArbitrarySize;16;NumericalValues;IDsAndValuesString;C1 C10 C11 C14 C18 C20 C21 C22 C5 CS F N11 N4 O10 O2 O9;5 1 1 1 14 4 2 1 2 2 1 1 1 1 3 1

```

```

FingerprintsVector;AtomTypesCount:SLogPAtomTypes:FixedSize;67;OrderedNumericalValues;IDsAndValuesString;C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 C13 C14 C15 C16 C17 C18 C19 C20 C21 C22 C23 C24 C25 C26 C27 CS N1 N2 N3 N4 N5 N6 N7 N8 N9 N10 N11 N12 N13 N14 NS O1 O2 O3 O4 O5 O6 O7 O8 O9 O10 O11 O12 OS F C1 Br I Hal P S1 S2 S3 Me1 Me2;5 0 0 0 2 0 0 0 0 1 1 0 0 1 0 0 0 14 0 4 2 1 0 0 0 0 0 2 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0...

```

```

FingerprintsVector;EStateIndices:ArbitrarySize;11;NumericalValues;IDsAndValuesString;SaaCH SaasC SaasN SdO SdssC SsCH3 SsF SsOH SssCH2 SssN

```

```
FingerprintsVector;MACCSKeyCount;322;OrderedNumericalValues;ValuesString;14 8 2 0 2 0 4 4 2 1 4 0 0 2 5 10 5 2 1 0 0 2 0 5 13 3 28 5 5 3 0 0  
0 4 2 1 1 0 1 1 0 0 2 1 0 0 0 0 0 0 0 0 0 0 0 0 22 5 3 0 0 0 1 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 11 0 2 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```



```
FingerprintsVector;TopologicalPharmacophoreAtomPairs:FixedSize:MinDistance:MaxDistance10;150;OrderedNumericalValues;ValuesString;18 0 0 1 0
0 0 2 0 0 1 0 0 0 0 22 12 8 0 0 1 2 0 0 0 0 0 0 0 18 6 3 1 0 0 0 1
0 0 1 0 0 0 0 22 13 6 0 0 5 7 0 0 2 0 0 0 0 0 28 9 5 1 0 0 0 1 0 0 1 0
0 0 0 36 16 10 0 0 3 4 0 0 1 0 0 0 0 0 37 10 8 0 0 0 0 1 0 0 0 0 0 0
0 35 10 9 0 0 3 3 0 0 1 0 0 0 0 0 28 7 7 4 0 0 0 0 0 0 0 0 0 0 18...
```

```
FingerprintsVector;TopologicalPharmacophoreAtomTriplets:ArbitrarySize:MinDistance:MaxDistance10;696;NumericalValues;IDsAndValuesString;Ar1-Ar1-Ar1-Ar1-H1 Ar1-Ar1-HBA1 Ar1-Ar1-HBD1 Ar1-H1-H1 Ar1-H1-HBA1 Ar1-H1-HBD1 Ar1-HBA1-HBD1 H1-H1-H1 H1-H1-HBA1 H1-H1-HBD1 H1-HBA1-HBA1 H1-HBA1-HBD1 H1-HBA1-NI1 H1-HBD1-NI1 HBA1-HBA1-NI1 HBA1-HBD1-NI1 Ar1-...;
46 106 8 3 83 11 4 1 21 5 3 1 2 2 1 1 1 100 101 18 11 145 132 26 14 23
28 3 3 5 4 61 45 10 4 16 20 7 5 1 3 4 5 3 1 1 1 5 4 2 1 2 2 2 1 1 1
119 123 24 15 185 202 41 25 22 17 3 5 85 95 18 11 23 17 3 1 1 6 4 ...
```

```
FingerprintsVector;TopologicalPharmacophoreAtomTriplets:FixedSize:MinDistance:MaxDistance10;2692;OrderedNumericalValues;ValuesString;46 106
8 3 0 0 83 11 4 0 0 0 1 0 0 0 0 0 0 0 21 5 3 0 0 1 2 2 0 0 1 0 0 0
0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 100 101 18 11 0 0 145 132 26
14 0 0 23 28 3 3 0 0 5 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 61 45 10 4 0
0 16 20 7 5 1 0 3 4 5 3 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 5 ...
```

OPTIONS

-a, --all

List all the available information.

--AverageBitDensity

List average bit density of fingerprint bit-vector strings.

--BitDensity

List bit density of fingerprints bit-vector strings data in each row.

--count

List number of data entries containing fingerprints bit-vector or vector strings data. This is default behavior.

-c, --ColMode *ColNum* | *ColLabel*

Specify how columns are identified in CSV/TSV *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 CSV/TSV *TextFile(s)* containing fingerprints data. Possible values: *col number* or *col label*. Default value: *first column containing the word Fingerprints in its column label*.

--FingerprintsField *FieldLabel*

Fingerprints field label to use during listing of fingerprints information for *SDFFile(s)*. Default value: *first data field label containing the word Fingerprints in its label*.

--FingerprintsType

List types of fingerprint strings: FingerprintsBitVector or FingerprintsVector.

--FingerprintsDescription

List types of fingerprints: PathLengthBits, PathLengthCount, MACCSKeyCount, ExtendedConnectivity and so on.

--FingerprintsSize

List size of fingerprints.

--FingerprintsBitStringFormat

List format of fingerprint bit-vector strings: BinaryString or HexadecimalString.

--FingerprintsBitOrder
List order of bits data in fingerprint bit-vector bit strings: Ascending or Descending.

--FingerprintsVectorValuesType
List type of values in fingerprint vector strings: OrderedNumericalValues, NumericalValues or AlphaNumericalValues.

--FingerprintsVectorValuesFormat
List format of values in fingerprint vector strings: ValuesString, IDsAndValuesString, IDsAndValuesPairsString, ValuesAndIDsString or ValuesAndIDsPairsString.

-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.

--NumOfOnBits
List number of on bits in fingerprints bit-vector strings data in each row.

--NumOfNonZeroValues
List number of non-zero values in fingerprints vector strings data in each row.

-w, --WorkingDir *DirName*
Location of working directory. Default: current directory.

EXAMPLES

To count number of lines containing fingerprints bit-vector or vector strings data present in FP file, in a column name containing Fingerprint substring in text file, and in a data field with Fingerprint substring in its label, type:

```
% InfoFingerprintsFiles.pl SampleFPBin.csv

% InfoFingerprintsFiles.pl SampleFPBin.sdf SampleFPBin.fpf
SampleFPBin.csv

% InfoFingerprintsFiles.pl SampleFPHex.sdf SampleFPHex.fpf
SampleFPHex.csv

% InfoFingerprintsFiles.pl SampleFPcount.sdf SampleFPcount.fpf
SampleFPcount.csv
```

To list all available information about fingerprints bit-vector or vector strings data present in FP file, in a column name containing Fingerprint substring in text file, and in a data field with Fingerprint substring in its label, type:

```
% InfoFingerprintsFiles.pl -a SampleFPHex.sdf SampleFPHex.fpf
SampleFPHex.csv

% InfoFingerprintsFiles.pl -a SampleFPcount.sdf SampleFPcount.fpf
SampleFPcount.csv
```

To list all available information about fingerprints bit-vector or vector strings data present in a column named Fingerprints in text file, type:

```
% InfoFingerprintsFiles.pl -a --ColMode ColLabel --FingerprintsCol
Fingerprints SampleFPHex.sdf

% InfoFingerprintsFiles.pl -a --ColMode ColLabel --FingerprintsCol
Fingerprints SampleFPcount.csv
```

To list all available information about fingerprints bit-vector or vector strings data present in a data field names Fingerprints in SD file, type:

```
% InfoFingerprintsFiles.pl -a --FingerprintsField Fingerprints
SampleFPHex.sdf

% InfoFingerprintsFiles.pl -a --FingerprintsField Fingerprints
SampleFPcount.sdf
```

To list bit density, average bit density, and number of on bits for fingerprints bit-vector strings data present in FP file, in a column name containing Fingerprint substring in text file, and in a data field with Fingerprint substring in its label, type:

```
% InfoFingerprintsFiles.pl --BitDensity --AverageBitDensity  
--NumOfOnBits SampleFPBin.csv SampleFPBin.sdf SampleFPBin.fpf
```

To list vector values type, format and number of non-zero values for fingerprints vector strings data present in FP file, in a column name containing Fingerprint substring in text file, and in a data field with Fingerprint substring in its label along with fingerprints type and description, type:

```
% InfoFingerprintsFiles.pl --FingerprintsType --FingerprintsDescription  
--FingerprintsVectorValuesType --FingerprintsVectorValuesFormat  
--NumOfNonZeroValues SampleFPcount.csv SampleFPcount.sdf  
SampleFPcount.fpf
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

SimilarityMatricesFingerprints.pl, SimilaritySearchingFingerprints.pl, AtomNeighborhoodsFingerprints.pl, AtomNeighborhoodsFingerprints.pl, ExtendedConnectivityFingerprints.pl, MACCSKeysFingerprints.pl, PathLengthFingerprints.pl, TopologicalAtomPairsFingerprints.pl, TopologicalAtomTorsionsFingerprints.pl, TopologicalPharmacophoreAtomPairsFingerprints.pl, TopologicalPharmacophoreAtomTripletsFingerprints.pl

COPYRIGHT

Copyright (C) 2015 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.