NAME

TopologicalAtomTripletsFingerprints.pl - Generate topological atom triplets fingerprints for SD files

SYNOPSIS

TopologicalAtomTripletsFingerprints.pl SDFile(s)...

TopologicalAtomTripletsFingerprints.pl [--AromaticityModel *AromaticityModelType*] [-a, --AtomIdentifierType *AtomicInvariantsAtomTypes*] [--AtomicInvariantsToUse *"AtomicInvariant,AtomicInvariant..."*] [--FunctionalClassesToUse *"FunctionalClass1,FunctionalClass2..."*] [--CompoundID *DataFieldName or LabelPrefixString*] [--CompoundIDLabel *text*] [--CompoundIDMode] [--DataFields *"FieldLabel1,FieldLabel2,..."*] [-d, --DataFieldsMode *All* | *Common* | *Specify* | *CompoundID*] [-f, --Filter *Yes* | *No*] [--FingerprintsLabel *text*] [-h, --help] [-k, --KeepLargestComponent *Yes* | *No*] [--MinDistance *number*] [--MaxDistance *number*] [--OutDelim *comma* | *tab* | *semicolon*] [--output *SD* | *FP* | *text* | *all*] [-o, --overwrite] [-q, --quote *Yes* | *No*] [-r, --root *RootName*] [-u, --UseTriangleInequality *Yes* | *No*] [-v, --VectorStringFormat *ValuesString, IDsAndValuesString* | *IDsAndValuesPairsString* | *ValuesAndIDsString* | *ValuesAndIDsPairsString*] [-w, --WorkingDir dirname] SDFile(s)...

DESCRIPTION

Generate topological atom triplets fingerprints for *SDFile(s)* and create appropriate SD, FP or CSV/TSV text file(s) containing fingerprints vector strings corresponding to molecular fingerprints.

Multiple SDFile names are separated by spaces. The valid file extensions are *.sdf* and *.sd*. All other file names are ignored. All the SD files in a current directory can be specified either by **.sdf* or the current directory name.

The current release of MayaChemTools supports generation of topological atom triplets fingerprints corresponding to following -a, --AtomIdentifierTypes:

```
AtomicInvariantsAtomTypes, DREIDINGAtomTypes, EStateAtomTypes,
FunctionalClassAtomTypes, MMFF94AtomTypes, SLogPAtomTypes,
SYBYLAtomTypes, TPSAAtomTypes, UFFAtomTypes
```

Based on the values specified for -a, --AtomIdentifierType and --AtomicInvariantsToUse, initial atom types are assigned to all non-hydrogen atoms in a molecule. Using the distance matrix for the molecule and initial atom types assigned to non-hydrogen atoms, all unique atom pairs within --MinDistance and --MaxDistance are identified and counted. An atom triplet identifier is generated for each unique atom triplet; the format of the atom triplet identifier is:

```
<ATx>-Dyz-<ATy>-Dxz-<ATz>-Dxy

ATx, ATy, ATz: Atom types assigned to atom x, atom y, and atom z
Dxy: Distance between atom x and atom y
Dxz: Distance between atom x and atom z
Dyz: Distance between atom y and atom z

where <AT1>-D23 <= <AT2>-D13 <= <AT3>-D12
```

The atom triplet identifiers for all unique atom triplets corresponding to non-hydrogen atoms constitute topological atom triplets fingerprints of the molecule.

Example of *SD* file containing topological atom triplets fingerprints string data:

```
... ...
... ...
$$$$
... ...
... ...
... ...
41 44  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
... ...
2  3  1  0  0  0  0
... ...
M  END
>  <CmpdID>
Cmpd1

>  <TopologicalAtomTripletsFingerprints>
```

```
FingerprintsVector;TopologicalAtomTriplets:AtomicInvariantsAtomTypes:Mi
nDistance1:MaxDistance10;3096;NumericalValues;IDsAndValuesString;C.X1.B
O1.H3-D1-C.X1.BO1.H3-D1-C.X3.BO3.H1-D2 C.X1.BO1.H3-D1-C.X2.BO2.H2-D10-C
.X3.BO4-D9 C.X1.BO1.H3-D1-C.X2.BO2.H2-D3-N.X3.BO3-D4 C.X1.BO1.H3-D1...;
1 2 2 2 2 2 2 2 8 8 4 8 4 4 2 2 2 2 4 2 2 2 4 2 2 2 2 1 2 2 4 4 4 2 2 2
4 4 4 8 4 4 2 4 4 4 2 4 4 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 8 8 ...

$$$$
... ...
... ...
```

Example of *FP* file containing topological atom triplets fingerprints string data:

```
#
# Package = MayaChemTools 7.4
# Release Date = Oct 21, 2010
#
# TimeStamp = Fri Mar 11 15:24:01 2011
#
# FingerprintsStringType = FingerprintsVector
#
# Description = TopologicalAtomTriplets:AtomicInvariantsAtomTypes:Mi...
# VectorStringFormat = IDsAndValuesString
# VectorValuesType = NumericalValues
#
Cmpd1 3096;C.X1.BO1.H3-D1-C.X1.BO1.H3-D1-C.X3.BO3.H1-D2...;1 2 2 2 2...
Cmpd2 1093;C.X1.BO1.H3-D1-C.X1.BO1.H3-D3-C.X2.BO2.H2-D4...;2 2 2 2 2...
... ...
... ..
```

Example of CSV *Text* file containing topological atom triplets fingerprints string data:

```
"CompoundID","TopologicalAtomTripletsFingerprints"
"Cmpd1","FingerprintsVector;TopologicalAtomTriplets:AtomicInvariantsAto
mTypes:MinDistance1:MaxDistance10;3096;NumericalValues;IDsAndValuesStri
ng;C.X1.BO1.H3-D1-C.X1.BO1.H3-D1-C.X3.BO3.H1-D2 C.X1.BO1.H3-D1-C.X2.BO2
.H2-D10-C.X3.BO4-D9 C.X1.BO1.H3-D1-C.X2.BO2.H2-D3-N.X3.BO3-D4 C.X1....;
1 2 2 2 2 2 2 2 8 8 4 8 4 4 2 2 2 2 4 2 2 2 4 2 2 2 2 1 2 2 4 4 4 2 2 2
4 4 4 8 4 4 2 4 4 4 2 4 4 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 8 8 ...
... ...
... ...
```

The current release of MayaChemTools generates the following types of topological atom triplets fingerprints vector strings:

```
FingerprintsVector;TopologicalAtomTriplets:AtomicInvariantsAtomTypes:M
inDistance1:MaxDistance10;3096;NumericalValues;IDsAndValuesString;C.X1
.BO1.H3-D1-C.X1.BO1.H3-D1-C.X3.BO3.H1-D2 C.X1.BO1.H3-D1-C.X2.BO2.H2-D1
0-C.X3.BO4-D9 C.X1.BO1.H3-D1-C.X2.BO2.H2-D3-N.X3.BO3-D4 C.X1.BO1.H3-D1
-C.X2.BO2.H2-D4-C.X2.BO2.H2-D5 C.X1.BO1.H3-D1-C.X2.BO2.H2-D6-C.X3....;
1 2 2 2 2 2 2 2 8 8 4 8 4 4 2 2 2 2 4 2 2 2 4 2 2 2 2 1 2 2 4 4 4 2 2
2 4 4 4 8 4 4 2 4 4 4 2 4 4 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 8...

FingerprintsVector;TopologicalAtomTriplets:AtomicInvariantsAtomTypes:M
inDistance1:MaxDistance10;3096;NumericalValues;IDsAndValuesPairsString
;C.X1.BO1.H3-D1-C.X1.BO1.H3-D1-C.X3.BO3.H1-D2 1 C.X1.BO1.H3-D1-C.X2.BO
2.H2-D10-C.X3.BO4-D9 2 C.X1.BO1.H3-D1-C.X2.BO2.H2-D3-N.X3.BO3-D4 2 C.X
1.BO1.H3-D1-C.X2.BO2.H2-D4-C.X2.BO2.H2-D5 2 C.X1.BO1.H3-D1-C.X2.BO2.H2
-D6-C.X3.BO3.H1-D5 2 C.X1.BO1.H3-D1-C.X2.BO2.H2-D6-C.X3.BO3.H1-D7 2...

FingerprintsVector;TopologicalAtomTriplets:DREIDINGAtomTypes:MinDistan
ce1:MaxDistance10;2377;NumericalValues;IDsAndValuesString;C_2-D1-C_2-D
9-C_3-D10 C_2-D1-C_2-D9-C_R-D10 C_2-D1-C_3-D1-C_3-D2 C_2-D1-C_3-D10-C_
3-D9 C_2-D1-C_3-D2-C_3-D3 C_2-D1-C_3-D2-C_R-D3 C_2-D1-C_3-D3-C_3-D4 C_
2-D1-C_3-D3-N_R-D4 C_2-D1-C_3-D3-O_3-D2 C_2-D1-C_3-D4-C_3-D5 C_2-D...;
1 1 1 2 1 1 3 1 1 2 2 1 1 1 1 1 1 1 2 1 3 4 5 1 1 6 4 2 2 3 1 1 1 2
2 1 2 1 1 2 2 2 1 2 1 2 1 1 3 3 2 6 4 2 1 1 1 2 2 1 1 1 1 1 1 1 1 1...
```

```
FingerprintsVector;TopologicalAtomTriplets:EStateAtomTypes:MinDistance
1:MaxDistance10;3298;NumericalValues;IDsAndValuesString;aaCH-D1-aaCH-D
1-aaCH-D2 aaCH-D1-aaCH-D1-aasC-D2 aaCH-D1-aaCH-D10-aaCH-D9 aaCH-D1-aaC
H-D10-aasC-D9 aaCH-D1-aaCH-D2-aaCH-D3 aaCH-D1-aaCH-D2-aasC-D1 aaCH-D1-
aaCH-D2-aasC-D3 aaCH-D1-aaCH-D3-aasC-D2 aaCH-D1-aaCH-D4-aasC-D5 aa...;
6 4 24 4 16 8 8 4 8 8 8 12 10 14 4 16 24 4 12 2 2 4 1 10 2 2 15 2 2 2
2 2 2 14 4 2 2 2 2 1 2 10 2 2 4 1 2 4 8 3 3 3 4 6 4 2 2 3 3 1 1 1 2 1
2 2 4 2 3 2 1 2 4 5 3 2 2 1 2 4 3 2 8 12 6 2 2 4 4 7 1 4 2 4 2 2 2 ...

FingerprintsVector;TopologicalAtomTriplets:FunctionalClassAtomTypes:Mi
nDistance1:MaxDistance10;2182;NumericalValues;IDsAndValuesString;Ar-D1
-Ar-D1-Ar-D2 Ar-D1-Ar-D1-Ar.HBA-D2 Ar-D1-Ar-D10-Ar-D9 Ar-D1-Ar-D10-Hal
-D9 Ar-D1-Ar-D2-Ar-D2 Ar-D1-Ar-D2-Ar-D3 Ar-D1-Ar-D2-Ar.HBA-D1 Ar-D1-Ar
-D2-Ar.HBA-D2 Ar-D1-Ar-D2-Ar.HBA-D3 Ar-D1-Ar-D2-HBD-D1 Ar-D1-Ar-D2...;
27 1 32 2 2 63 3 2 1 2 1 2 3 1 1 40 3 1 2 2 2 2 4 2 2 47 4 2 2 1 2 1 5
2 2 51 4 3 1 3 1 9 1 1 50 3 3 4 1 9 50 2 2 3 3 5 45 1 1 1 2 1 2 2 3 3
4 4 3 2 1 1 3 4 5 5 3 1 2 3 2 3 5 7 2 7 3 7 1 1 2 2 2 2 3 1 4 3 1 2...

FingerprintsVector;TopologicalAtomTriplets:MMFF94AtomTypes:MinDistance
1:MaxDistance10;2966;NumericalValues;IDsAndValuesString;C5A-D1-C5A-D1-
N5-D2 C5A-D1-C5A-D2-C5B-D2 C5A-D1-C5A-D3-CB-D2 C5A-D1-C5A-D3-CR-D2 C5A
-D1-C5B-D1-C5B-D2 C5A-D1-C5B-D2-C=ON-D1 C5A-D1-C5B-D2-CB-D1 C5A-D1-C5B
-D3-C=ON-D2 C5A-D1-C5B-D3-CB-D2 C5A-D1-C=ON-D3-NC=O-D2 C5A-D1-C=ON-D3-
O=CN-D2 C5A-D1-C=ON-D4-NC=O-D3 C5A-D1-C=ON-D4-O=CN-D3 C5A-D1-CB-D1-...

FingerprintsVector;TopologicalAtomTriplets:SLogPAtomTypes:MinDistance1
:MaxDistance10;3710;NumericalValues;IDsAndValuesString;C1-D1-C1-D1-C11
-D2 C1-D1-C1-D1-CS-D2 C1-D1-C1-D10-C5-D9 C1-D1-C1-D3-C10-D2 C1-D1-C1-D
3-C5-D2 C1-D1-C1-D3-CS-D2 C1-D1-C1-D3-CS-D4 C1-D1-C1-D4-C10-D5 C1-D1-C
1-D4-C11-D5 C1-D1-C1-D5-C10-D4 C1-D1-C1-D5-C5-D4 C1-D1-C1-D6-C11-D7 C1
-D1-C1-D6-CS-D5 C1-D1-C1-D6-CS-D7 C1-D1-C1-D8-C11-D9 C1-D1-C1-D8-CS...

FingerprintsVector;TopologicalAtomTriplets:SYBYLAtomTypes:MinDistance1
:MaxDistance10;2332;NumericalValues;IDsAndValuesString;C.2-D1-C.2-D9-C
.3-D10 C.2-D1-C.2-D9-C.ar-D10 C.2-D1-C.3-D1-C.3-D2 C.2-D1-C.3-D10-C.3-
D9 C.2-D1-C.3-D2-C.3-D3 C.2-D1-C.3-D2-C.ar-D3 C.2-D1-C.3-D3-C.3-D4 C.2
-D1-C.3-D3-N.ar-D4 C.2-D1-C.3-D3-O.3-D2 C.2-D1-C.3-D4-C.3-D5 C.2-D1-C.
3-D5-C.3-D6 C.2-D1-C.3-D5-O.3-D4 C.2-D1-C.3-D6-C.3-D7 C.2-D1-C.3-D7...

FingerprintsVector;TopologicalAtomTriplets:TPSAAtomTypes:MinDistance1:
MaxDistance10;1007;NumericalValues;IDsAndValuesString;N21-D1-N7-D3-Non
e-D4 N21-D1-N7-D5-None-D4 N21-D1-None-D1-None-D2 N21-D1-None-D2-None-D
2 N21-D1-None-D2-None-D3 N21-D1-None-D3-None-D4 N21-D1-None-D4-None-D5
 N21-D1-None-D4-O3-D3 N21-D1-None-D4-O4-D3 N21-D1-None-D5-None-D6 N21-
D1-None-D6-None-D7 N21-D1-None-D6-O4-D5 N21-D1-None-D7-None-D8 N21-...

FingerprintsVector;TopologicalAtomTriplets:UFFAtomTypes:MinDistance1:M
axDistance10;2377;NumericalValues;IDsAndValuesString;C_2-D1-C_2-D9-C_3
-D10 C_2-D1-C_2-D9-C_R-D10 C_2-D1-C_3-D1-C_3-D2 C_2-D1-C_3-D10-C_3-D9
C_2-D1-C_3-D2-C_3-D3 C_2-D1-C_3-D2-C_R-D3 C_2-D1-C_3-D3-C_3-D4 C_2-D1-
C_3-D3-N_R-D4 C_2-D1-C_3-D3-O_3-D2 C_2-D1-C_3-D4-C_3-D5 C_2-D1-C_3-D5-
C_3-D6 C_2-D1-C_3-D5-O_3-D4 C_2-D1-C_3-D6-C_3-D7 C_2-D1-C_3-D7-C_3-...
```

OPTIONS

--AromaticityModel *MDLAromaticityModel* | *TriposAromaticityModel* | *MMFFAromaticityModel* | *ChemAxonBasicAromaticityModel* | *ChemAxonGeneralAromaticityModel* | *DaylightAromaticityModel* | *MayaChemToolsAromaticityModel*

Specify aromaticity model to use during detection of aromaticity. Possible values in the current release are: *MDLAromaticityModel, TriposAromaticityModel, MMFFAromaticityModel, ChemAxonBasicAromaticityModel, ChemAxonGeneralAromaticityModel, DaylightAromaticityModel or MayaChemToolsAromaticityModel*. Default value: *MayaChemToolsAromaticityModel*.

The supported aromaticity model names along with model specific control parameters are defined in AromaticityModelsData.csv, which is distributed with the current release and is available under lib/data directory. Molecule.pm module retrieves data from this file during class instantiation and makes it available to method DetectAromaticity for detecting aromaticity corresponding to a specific

-a, --AtomIdentifierType *AtomicInvariantsAtomTypes | DREIDINGAtomTypes | EStateAtomTypes | FunctionalClassAtomTypes | MMFF94AtomTypes | SLogPAtomTypes | SYBYLAtomTypes | TPSAAtomTypes | UFFAtomTypes*

Specify atom identifier type to use for assignment of initial atom identifier to non-hydrogen atoms during calculation of topological atom triplets fingerprints. Possible values in the current release are: *AtomicInvariantsAtomTypes, DREIDINGAtomTypes, EStateAtomTypes, FunctionalClassAtomTypes, MMFF94AtomTypes, SLogPAtomTypes, SYBYLAtomTypes, TPSAAtomTypes, UFFAtomTypes*. Default value: *AtomicInvariantsAtomTypes*.

--AtomicInvariantsToUse *"AtomicInvariant,AtomicInvariant..."*

This value is used during *AtomicInvariantsAtomTypes* value of a, --AtomIdentifierType option. It's a list of comma separated valid atomic invariant atom types.

Possible values for atomic invariants are: *AS, X, BO, LBO, SB, DB, TB, H, Ar, RA, FC, MN, SM*. Default value: *AS,X,BO,H,FC*.

The atomic invariants abbreviations correspond to:

```
AS = Atom symbol corresponding to element symbol

X<n>   = Number of non-hydrogen atom neighbors or heavy atoms
BO<n> = Sum of bond orders to non-hydrogen atom neighbors or heavy atoms
LBO<n> = Largest bond order of non-hydrogen atom neighbors or heavy atoms
SB<n> = Number of single bonds to non-hydrogen atom neighbors or heavy atoms
DB<n> = Number of double bonds to non-hydrogen atom neighbors or heavy atoms
TB<n> = Number of triple bonds to non-hydrogen atom neighbors or heavy atoms
H<n>   = Number of implicit and explicit hydrogens for atom
Ar     = Aromatic annotation indicating whether atom is aromatic
RA     = Ring atom annotation indicating whether atom is a ring
FC<+n/-n> = Formal charge assigned to atom
MN<n> = Mass number indicating isotope other than most abundant isotope
SM<n> = Spin multiplicity of atom. Possible values: 1 (singlet), 2 (doublet) or
        3 (triplet)
```

Atom type generated by AtomTypes::AtomicInvariantsAtomTypes class corresponds to:

```
AS.X<n>.BO<n>.LBO<n>.<SB><n>.<DB><n>.<TB><n>.H<n>.Ar.RA.FC<+n/-n>.MN<n>.SM<n>
```

Except for AS which is a required atomic invariant in atom types, all other atomic invariants are optional. Atom type specification doesn't include atomic invariants with zero or undefined values.

In addition to usage of abbreviations for specifying atomic invariants, the following descriptive words are also allowed:

```
X : NumOfNonHydrogenAtomNeighbors or NumOfHeavyAtomNeighbors
BO : SumOfBondOrdersToNonHydrogenAtoms or SumOfBondOrdersToHeavyAtoms
LBO : LargestBondOrderToNonHydrogenAtoms or LargestBondOrderToHeavyAtoms
SB :  NumOfSingleBondsToNonHydrogenAtoms or NumOfSingleBondsToHeavyAtoms
DB : NumOfDoubleBondsToNonHydrogenAtoms or NumOfDoubleBondsToHeavyAtoms
TB : NumOfTripleBondsToNonHydrogenAtoms or NumOfTripleBondsToHeavyAtoms
H :  NumOfImplicitAndExplicitHydrogens
Ar : Aromatic
RA : RingAtom
FC : FormalCharge
MN : MassNumber
SM : SpinMultiplicity
```

AtomTypes::AtomicInvariantsAtomTypes module is used to assign atomic invariant atom types.

--FunctionalClassesToUse *"FunctionalClass1,FunctionalClass2..."*

This value is used during *FunctionalClassAtomTypes* value of a, --AtomIdentifierType option. It's a list of comma separated valid functional classes.

Possible values for atom functional classes are: *Ar, CA, H, HBA, HBD, Hal, NI, PI, RA*. Default value [Ref 24]: *HBD,HBA,PI,NI,Ar,Hal*.

The functional class abbreviations correspond to:

```
HBD: HydrogenBondDonor
HBA: HydrogenBondAcceptor
PI :  PositivelyIonizable
NI : NegativelyIonizable
Ar : Aromatic
Hal : Halogen
```

```
H : Hydrophobic
RA : RingAtom
CA : ChainAtom
```

```
Functional class atom type specification for an atom corresponds to:

Ar.CA.H.HBA.HBD.Hal.NI.PI.RA
```

AtomTypes::FunctionalClassAtomTypes module is used to assign functional class atom types. It uses following definitions [Ref 60-61, Ref 65-66]:

```
HydrogenBondDonor: NH, NH2, OH
HydrogenBondAcceptor: N[!H], O
PositivelyIonizable: +, NH2
NegativelyIonizable: -, C(=O)OH, S(=O)OH, P(=O)OH
```

--CompoundID *DataFieldName or LabelPrefixString*

This value is --CompoundIDMode specific and indicates how compound ID is generated.

For *DataField* value of --CompoundIDMode option, it corresponds to datafield label name whose value is used as compound ID; otherwise, it's a prefix string used for generating compound IDs like LabelPrefixString<Number>. Default value, *Cmpd*, generates compound IDs which look like Cmpd<Number>.

Examples for *DataField* value of --CompoundIDMode:

```
MolID
ExtReg
```

Examples for *LabelPrefix* or *MolNameOrLabelPrefix* value of --CompoundIDMode:

```
Compound
```

The value specified above generates compound IDs which correspond to Compound<Number> instead of default value of Cmpd<Number>.

--CompoundIDLabel *text*

Specify compound ID column label for CSV/TSV text file(s) used during *CompoundID* value of --DataFieldsMode option. Default value: *CompoundID*.

--CompoundIDMode *DataField* | *MolName* | *LabelPrefix* | *MolNameOrLabelPrefix*

Specify how to generate compound IDs and write to FP or CSV/TSV text file(s) along with generated fingerprints for *FP* | *text* | *all* values of --output option: use a *SDFile(s)* datafield value; use molname line from *SDFile(s)*; generate a sequential ID with specific prefix; use combination of both MolName and LabelPrefix with usage of LabelPrefix values for empty molname lines.

Possible values: *DataField* | *MolName* | *LabelPrefix* | *MolNameOrLabelPrefix*. Default value: *LabelPrefix*.

For *MolNameAndLabelPrefix* value of --CompoundIDMode, molname line in *SDFile(s)* takes precedence over sequential compound IDs generated using *LabelPrefix* and only empty molname values are replaced with sequential compound IDs.

This is only used for *CompoundID* value of --DataFieldsMode option.

--DataFields *"FieldLabel1,FieldLabel2,..."*

Comma delimited list of *SDFiles(s)* data fields to extract and write to CSV/TSV text file(s) along with generated fingerprints for *text* | *all* values of --output option.

This is only used for *Specify* value of --DataFieldsMode option.

Examples:

```
Extreg
MolID,CompoundName
```

-d, --DataFieldsMode *All* | *Common* | *Specify* | *CompoundID*

Specify how data fields in *SDFile(s)* are transferred to output CSV/TSV text file(s) along with generated fingerprints for *text* | *all* values of --output option: transfer all SD data field; transfer SD data files common to all compounds; extract specified data fields; generate a compound ID using molname line, a compound prefix, or a combination of both. Possible values: *All* | *Common* | *specify* | *CompoundID*. Default value: *CompoundID*.

-f, --Filter *Yes* | *No*

Specify whether to check and filter compound data in SDFile(s). Possible values: *Yes or No*. Default value: *Yes*.

By default, compound data is checked before calculating fingerprints and compounds containing atom data corresponding to non-element symbols or no atom data are ignored.

--FingerprintsLabel *text*

SD data label or text file column label to use for fingerprints string in output SD or CSV/TSV text file(s) specified by --output. Default value: *TopologicalAtomTripletsFingerprints*.

-h, --help

Print this help message.

-k, --KeepLargestComponent *Yes* | *No*

Generate fingerprints for only the largest component in molecule. Possible values: *Yes or No*. Default value: *Yes*.

For molecules containing multiple connected components, fingerprints can be generated in two different ways: use all connected components or just the largest connected component. By default, all atoms except for the largest connected component are deleted before generation of fingerprints.

--MinDistance *number*

Minimum bond distance between atom triplets for generating topological atom triplets. Default value: *1*. Valid values: positive integers and less than --MaxDistance.

--MaxDistance *number*

Maximum bond distance between atom triplets for generating topological atom triplets. Default value: *10*. Valid values: positive integers and greater than --MinDistance.

--OutDelim *comma* | *tab* | *semicolon*

Delimiter for output CSV/TSV text file(s). Possible values: *comma, tab, or semicolon* Default value: *comma*

--output *SD* | *FP* | *text* | *all*

Type of output files to generate. Possible values: *SD, FP, text, or all*. Default value: *text*.

-o, --overwrite

Overwrite existing files.

-q, --quote *Yes* | *No*

Put quote around column values in output CSV/TSV text file(s). Possible values: *Yes or No*. Default value: *Yes*.

-r, --root *RootName*

New file name is generated using the root: <Root>.<Ext>. Default for new file names: <SDFileName><TopologicalAtomTripletsFP>.<Ext>. The file type determines <Ext> value. The sdf, fpf, csv, and tsv <Ext> values are used for SD, FP, comma/semicolon, and tab delimited text files, respectively.This option is ignored for multiple input files.

-u, --UseTriangleInequality *Yes* | *No*

Specify whether to imply triangle distance inequality test to distances between atom pairs in atom triplets during generation of atom triplets generation. Possible values: *Yes or No*. Default value: *No*.

Triangle distance inequality test implies that distance or binned distance between any two atom pairs in an atom triplet must be less than the sum of distances or binned distances between other two atoms pairs and greater than the difference of their distances.

```
For atom triplet ATx-Dyz-ATy-Dxz-ATz-Dxy to satisfy triangle inequality:

Dyz > |Dxz - Dxy| and Dyz < Dxz + Dxy
Dxz > |Dyz - Dxy| and Dyz < Dyz + Dxy
Dxy > |Dyz - Dxz| and Dxy < Dyz + Dxz
```

-v, --VectorStringFormat *IDsAndValuesString* | *IDsAndValuesPairsString* | *ValuesAndIDsString* | *ValuesAndIDsPairsString*

Format of fingerprints vector string data in output SD, FP or CSV/TSV text file(s) specified by --output option. Possible values: *IDsAndValuesString* | *IDsAndValuesPairsString* | *ValuesAndIDsString* | *ValuesAndIDsPairsString*. Default value: *IDsAndValuesString*.

Examples:

```
FingerprintsVector;TopologicalAtomTriplets:AtomicInvariantsAtomTypes:M
inDistance1:MaxDistance10;3096;NumericalValues;IDsAndValuesString;C.X1
.BO1.H3-D1-C.X1.BO1.H3-D1-C.X3.BO3.H1-D2 C.X1.BO1.H3-D1-C.X2.BO2.H2-D1
0-C.X3.BO4-D9 C.X1.BO1.H3-D1-C.X2.BO2.H2-D3-N.X3.BO3-D4 C.X1.BO1.H3-D1
-C.X2.BO2.H2-D4-C.X2.BO2.H2-D5 C.X1.BO1.H3-D1-C.X2.BO2.H2-D6-C.X3....;
1 2 2 2 2 2 2 2 8 8 4 8 4 4 2 2 2 2 4 2 2 2 4 2 2 2 2 1 2 2 4 4 4 2 2
2 4 4 4 8 4 4 2 4 4 4 2 4 4 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 8...

FingerprintsVector;TopologicalAtomTriplets:AtomicInvariantsAtomTypes:M
inDistance1:MaxDistance10;3096;NumericalValues;IDsAndValuesPairsString
;C.X1.BO1.H3-D1-C.X1.BO1.H3-D1-C.X3.BO3.H1-D2 1 C.X1.BO1.H3-D1-C.X2.BO
2.H2-D10-C.X3.BO4-D9 2 C.X1.BO1.H3-D1-C.X2.BO2.H2-D3-N.X3.BO3-D4 2 C.X
1.BO1.H3-D1-C.X2.BO2.H2-D4-C.X2.BO2.H2-D5 2 C.X1.BO1.H3-D1-C.X2.BO2.H2
-D6-C.X3.BO3.H1-D5 2 C.X1.BO1.H3-D1-C.X2.BO2.H2-D6-C.X3.BO3.H1-D7 2...
```

-w, --WorkingDir *DirName*

Location of working directory. Default value: current directory.

EXAMPLES

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using atomic invariants atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using atomic invariants atom types in IDsAndValuesString format and create SampleTATFP.sdf, SampleTATFP.fpf and SampleTATFP.csv files containing sequential compound IDs in CSV file along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl --output all -r SampleTATFP
  -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using atomic invariants atom types in IDsAndValuesPairsString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl --VectorStringFormat
  IDsAndValuesPairsString  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using DREIDING atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a DREIDINGAtomTypes
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using E-state atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a EStateAtomTypes
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using functional class atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a FunctionalClassAtomTypes
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using DREIDING atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a DREIDINGAtomTypes
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using MM94 atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a MMFF94AtomTypes
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using SLogP atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a SLogPAtomTypes
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using SYBYL atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a SYBYLAtomTypes
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using TPSA atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a TPSAAtomTypes
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using UFF atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a UFFAtomTypes
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 6 using atomic invariants atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a AtomicInvariantsAtomTypes
 --MinDistance 1 --MaxDistance 6 -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using only AS,X atomic invariants atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing sequential compound IDs along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a AtomicInvariantsAtomTypes
  --AtomicInvariantsToUse "AS,X" --MinDistance 1 --MaxDistance 6
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using atomic invariants atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing compound ID from molecule name line along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a AtomicInvariantsAtomTypes
  --DataFieldsMode CompoundID -CompoundIDMode MolName
  -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using atomic invariants atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing compound IDs using specified data field along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a AtomicInvariantsAtomTypes
  --DataFieldsMode CompoundID -CompoundIDMode DataField --CompoundID
  Mol_ID -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using

atomic invariants atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing compound ID using combination of molecule name line and an explicit compound prefix along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a AtomicInvariantsAtomTypes
  --DataFieldsMode CompoundID -CompoundIDMode MolnameOrLabelPrefix
  --CompoundID Cmpd --CompoundIDLabel MolID -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using atomic invariants atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing specific data fields columns along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a AtomicInvariantsAtomTypes
  --DataFieldsMode Specify --DataFields Mol_ID -r SampleTATFP
  -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using atomic invariants atom types in IDsAndValuesString format and create a SampleTATFP.csv file containing common data fields columns along with fingerprints vector strings data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a AtomicInvariantsAtomTypes
  --DataFieldsMode Common -r SampleTATFP -o Sample.sdf
```

To generate topological atom triplets fingerprints corresponding to bond distances from 1 through 10 using atomic invariants atom types in IDsAndValuesString format and create SampleTATFP.sdf, SampleTATFP.fpf and SampleTATFP.csv files containing all data fields columns in CSV file along with fingerprints data, type:

```
% TopologicalAtomTripletsFingerprints.pl -a AtomicInvariantsAtomTypes
  --DataFieldsMode All  --output all -r SampleTATFP
  -o Sample.sdf
```

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

InfoFingerprintsFiles.pl, SimilarityMatricesFingerprints.pl, AtomNeighborhoodsFingerprints.pl, ExtendedConnectivityFingerprints.pl, MACCSKeysFingerprints.pl, PathLengthFingerprints.pl, TopologicalAtomTorsionsFingerprints.pl, TopologicalPharmacophoreAtomPairsFingerprints.pl, TopologicalPharmacophoreAtomTripletsFingerprints.pl

COPYRIGHT

Copyright (C) 2019 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.