NAME

FingerprintsStringUtil

SYNOPSIS

use Fingerprints::FingerprintsStringUtil;

use Fingerprints::FingerprintsStringUtil qw(:all);

DESCRIPTION

FingerprintsStringUtil module provides the following functions:

AreFingerprintsStringValuesValid, GenerateFingerprintsBitVectorString, GenerateFingerprintsString, GenerateFingerprintsVectorString, GetDefaultBitStringFormat, GetDefaultBitsOrder, GetDefaultVectorStringFormat, GetFingeprintsStringDelimiter, GetFingerprintsStringTypeAndDescription, GetFingerprintsStringValues, ParseFingerprintsBitVectorString, ParseFingerprintsString, ParseFingerprintsVectorString

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

```
    FingerprintsVector;AtomNeighborhoods:AtomicInvariantsAtomTypes:MinRadi
    us0:MaxRadius2;41;AlphaNumericalValues;ValuesString;NR0-C.X1.BO1.H3-AT
    C1:NR1-C.X3.BO3.H1-ATC1:NR2-C.X1.BO1.H3-ATC1:NR2-C.X3.BO4-ATC1 NR0-C.X
    1.BO1.H3-ATC1:NR1-C.X3.BO3.H1-ATC1:NR2-C.X1.BO1.H3-ATC1:NR2-C.X3.BO4-A
    TC1 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:ArbitraryS
    ize;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;Nume
    ricalValues;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;OrderedN
    umericalValues;IDsAndValuesString;C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C
    12 C13 C14 C15 C16 C17 C18 C19 C20 C21 C22 C23 C24 C25 C26 C27 CS N1 N
    2 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 Cl 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 0...

    FingerprintsVector;EStateIndicies:ArbitrarySize;11;NumericalValues;IDs
    AndValuesString;SaaCH SaasC SaasN SdO SdssC SsCH3 SsF SsOH SssCH2 SssN
    H SsssCH;24.778 4.387 1.993 25.023 -1.435 3.975 14.006 29.759 -0.073 3
    .024 -2.270

    FingerprintsVector;EStateIndicies:FixedSize;87;OrderedNumericalValues;
    ValuesString;0 0 0 0 0 0 0 3.975 0 -0.073 0 0 24.778 -2.270 0 0 -1.435
    4.387 0 0 0 0 0 0 3.024 0 0 0 0 0 0 0 1.993 0 29.759 25.023 0 0 0 0 1
    4.006 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

    FingerprintsVector;ExtendedConnectivity:AtomicInvariantsAtomTypes:Radi
    us2;60;AlphaNumericalValues;ValuesString;73555770 333564680 352413391
    666191900 1001270906 1371674323 1481469939 1977749791 2006158649 21414
    08799 49532520 64643108 79385615 96062769 273726379 564565671 85514103
    5 906706094 988546669 1018231313 1032696425 1197507444 1331250018 1338
    532734 1455473691 1607485225 1609687129 1631614296 1670251330 17303...

    FingerprintsVector;ExtendedConnectivityCount:AtomicInvariantsAtomTypes
    :Radius2;60;NumericalValues;IDsAndValuesString;73555770 333564680 3524
    13391 666191900 1001270906 1371674323 1481469939 1977749791 2006158649
    2141408799 49532520 64643108 79385615 96062769 273726379 564565671...;
    3 2 1 1 14 1 2 10 4 3 1 1 1 1 2 1 2 1 1 1 2 3 1 1 2 1 3 3 8 2 2 2 6 2
    1 2 1 1 2 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 1 1 2 1 1

    FingerprintsBitVector;ExtendedConnectivityBits:AtomicInvariantsAtomTyp
    es:Radius2;1024;BinaryString;Ascending;0000000000000000000000000000100
    0000000001010000000110000011000000000000100000000000000000000000001000001
    1000000110000000000000000000000000001001100000000000000000000000000010000
    0000000000000000000000000010000000000000000001000000000000000000000000
    0000000000010000100001000000000000101000000000000000100000000000000...

    FingerprintsVector;ExtendedConnectivity:FunctionalClassAtomTypes:Radiu
```

```
s2;57;AlphaNumericalValues;ValuesString;24769214 508787397 850393286 8
62102353 981185303 1231636850 1649386610 1941540674 263599683 32920567
1 571109041 639579325 683993318 723853089 810600886 885767127 90326012
7 958841485 981022393 1126908698 1152248391 1317567065 1421489994 1455
632544 1557272891 1826413669 1983319256 2015750777 2029559552 20404...

FingerprintsVector;ExtendedConnectivity:EStateAtomTypes:Radius2;62;Alp
haNumericalValues;ValuesString;25189973 528584866 662581668 671034184
926543080 1347067490 1738510057 1759600920 2034425745 2097234755 21450
44754 96779665 180364292 341712110 345278822 386540408 387387308 50430
1706 617094135 771528807 957666640 997798220 1158349170 1291258082 134
1138533 1395329837 1420277211 1479584608 1486476397 1487556246 1566...

FingerprintsBitVector;MACCSKeyBits;166;BinaryString;Ascending;00000000
0000000000000000000000000000000001001000010010000000010010000000011100
0100101010111100011011000100110110000011011110100110111111111111011111
11111111111110111000

FingerprintsBitVector;MACCSKeyBits;322;BinaryString;Ascending;11101011
1110011111100101111111000111101100110000000000011100010000000000000000
0000000000000000000000000000000000000000000000101000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000011000000000000000000000000000000
0000000000000000000000000000000000000000

FingerprintsVector;MACCSKeyCount;166;OrderedNumericalValues;ValuesStri
ng;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 1 0 0 3 0 0 0 0 4 0 0 2 0 0 0 0 0 0 0 0 2 0 0 2 0 0 0 0
0 0 0 0 1 1 8 0 0 0 1 0 0 1 0 1 0 1 0 3 1 3 1 0 0 0 1 2 0 11 1 0 0 0
5 0 0 1 2 0 1 1 0 0 0 0 0 1 1 0 1 1 1 1 0 4 0 0 1 1 0 4 6 1 1 1 2 1 1
3 5 2 2 0 5 3 5 1 1 2 5 1 2 1 2 4 8 3 5 5 2 2 0 3 5 4 1

FingerprintsVector;MACCSKeyCount;322;OrderedNumericalValues;ValuesStri
ng;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 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 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 0 0 0 0 0 ...

FingerprintsBitVector;PathLengthBits:AtomicInvariantsAtomTypes:MinLeng
th1:MaxLength8;1024;BinaryString;Ascending;001000010011010101011000110
0100010101011000101001011100110001000001000100110100001001001001001000
0010110100000011100100100000001010010010000000000110000000101001011100
0010000000100010101010000001001111001101110110110000000101101110011101
0101100011000000001000100000110000101000111011000010000001000100000...

FingerprintsVector;PathLengthCount:AtomicInvariantsAtomTypes:MinLength
1:MaxLength8;432;NumericalValues;IDsAndValuesPairsString;C.X1.BO1.H3 2
C.X2.BO2.H2 4 C.X2.BO3.H1 14 C.X3.BO3.H1 3 C.X3.BO4 10 F.X1.BO1 1 N.X
2.BO2.H1 1 N.X3.BO3 1 O.X1.BO1.H1 3 O.X1.BO2 2 C.X1.BO1.H3C.X3.BO3.H1
2 C.X2.BO2.H2C.X2.BO2.H2 1 C.X2.BO2.H2C.X3.BO3.H1 4 C.X2.BO2.H2C.X3.BO
4 1 C.X2.BO2.H2N.X3.BO3 1 C.X2.BO3.H1:C.X2.BO3.H1 10 C.X2.BO3.H1:C....

FingerprintsVector;PathLengthCount:MMFF94AtomTypes:MinLength1:MaxLengt
h8;463;NumericalValues;IDsAndValuesPairsString;C5A 2 C5B 2 C=ON 1 CB 1
8 COO 1 CR 9 F 1 N5 1 NC=O 1 O=CN 1 O=CO 1 OC=O 1 OR 2 C5A:C5B 2 C5A:N
5 2 C5ACB 1 C5ACR 1 C5B:C5B 1 C5BC=ON 1 C5BCB 1 C=ON=O=CN 1 C=ONNC=O 1
CB:CB 18 CBF 1 CBNC=O 1 COO=O=CO 1 COOCR 1 COOOC=O 1 CRCR 7 CRN5 1 CR
OR 2 C5A:C5B:C5B 2 C5A:C5BC=ON 1 C5A:C5BCB 1 C5A:N5:C5A 1 C5A:N5CR ...

FingerprintsVector;TopologicalAtomPairs:AtomicInvariantsAtomTypes:MinD
istance1:MaxDistance10;223;NumericalValues;IDsAndValuesString;C.X1.BO1
.H3-D1-C.X3.BO3.H1 C.X2.BO2.H2-D1-C.X2.BO2.H2 C.X2.BO2.H2-D1-C.X3.BO3.
H1 C.X2.BO2.H2-D1-C.X3.BO4 C.X2.BO2.H2-D1-N.X3.BO3 C.X2.BO3.H1-D1-...;
2 1 4 1 1 10 8 1 2 6 1 2 2 1 2 1 2 2 1 2 1 5 1 10 12 2 2 1 2 1 9 1 3 1
1 1 2 2 1 3 6 1 6 14 2 2 2 3 1 3 1 8 2 2 1 3 2 6 1 2 2 5 1 3 1 23 1...

FingerprintsVector;TopologicalAtomPairs:FunctionalClassAtomTypes:MinDi
stance1:MaxDistance10;144;NumericalValues;IDsAndValuesString;Ar-D1-Ar
Ar-D1-Ar.HBA Ar-D1-HBD Ar-D1-Hal Ar-D1-None Ar.HBA-D1-None HBA-D1-NI H
BA-D1-None HBA.HBD-D1-NI HBA.HBD-D1-None HBD-D1-None NI-D1-None No...;
23 2 1 1 2 1 1 1 1 2 1 1 7 28 3 1 3 2 8 2 1 1 1 5 1 5 24 3 3 4 2 13 4
1 1 4 1 5 22 4 4 3 1 19 1 1 1 1 1 2 2 3 1 1 8 25 4 5 2 3 1 26 1 4 1 ...
```

```
FingerprintsVector;TopologicalAtomTorsions:AtomicInvariantsAtomTypes;3
3;NumericalValues;IDsAndValuesString;C.X1.BO1.H3-C.X3.BO3.H1-C.X3.BO4-
C.X3.BO4 C.X1.BO1.H3-C.X3.BO3.H1-C.X3.BO4-N.X3.BO3 C.X2.BO2.H2-C.X2.BO
2.H2-C.X3.BO3.H1-C.X2.BO2.H2 C.X2.BO2.H2-C.X2.BO2.H2-C.X3.BO3.H1-O...;
2 2 1 1 2 2 1 1 3 4 4 8 4 2 2 6 2 2 1 2 1 1 2 1 1 2 6 2 4 2 1 3 1

FingerprintsVector;TopologicalAtomTorsions:EStateAtomTypes;36;Numerica
lValues;IDsAndValuesString;aaCH-aaCH-aaCH-aaCH aaCH-aaCH-aaCH-aasC aaC
H-aaCH-aasC-aaCH aaCH-aaCH-aasC-aasC aaCH-aaCH-aasC-sF aaCH-aaCH-aasC-
ssNH aaCH-aasC-aasC-aasC aaCH-aasC-aasC-aasN aaCH-aasC-ssNH-dssC a...;
4 4 8 4 2 2 6 2 2 2 4 3 2 1 3 3 2 2 2 1 2 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2

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: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;TopologicalPharmacophoreAtomPairs:ArbitrarySize:Min
Distance1:MaxDistance10;54;NumericalValues;IDsAndValuesString;H-D1-H H
-D1-NI HBA-D1-NI HBD-D1-NI H-D2-H H-D2-HBA H-D2-HBD HBA-D2-HBA HBA-D2-
HBD H-D3-H H-D3-HBA H-D3-HBD H-D3-NI HBA-D3-NI HBD-D3-NI H-D4-H H-D4-H
BA H-D4-HBD HBA-D4-HBA HBA-D4-HBD HBD-D4-HBD H-D5-H H-D5-HBA H-D5-...;
18 1 2 1 22 12 8 1 2 18 6 3 1 1 1 22 13 6 5 7 2 28 9 5 1 1 1 36 16 10
3 4 1 37 10 8 1 35 10 9 3 3 1 28 7 7 4 18 16 12 5 1 2 1

FingerprintsVector;TopologicalPharmacophoreAtomPairs:FixedSize:MinDist
ance1: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 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 0 18...

FingerprintsVector;TopologicalPharmacophoreAtomTriplets:ArbitrarySize:
MinDistance1:MaxDistance10;696;NumericalValues;IDsAndValuesString;Ar1-
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 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:MinD
istance1:MaxDistance10;2692;OrderedNumericalValues;ValuesString;46 106
8 3 0 0 83 11 4 0 0 0 1 0 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 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 ...
```

FUNCTIONS

AreFingerprintsStringValuesValid

```
$Status = AreFingerprintsStringValuesValid($FPString);
```

Returns 0 or 1 based on whether *FingerprintsString* contains valid values.

GetDefaultBitStringFormat

```
$BitStringFormat = GetDefaultBitStringFormat();
```

Returns default BitStringFormat for fingerprints bit-vector strings.

GetDefaultBitsOrder

```
$BitsOrder = GetDefaultBitsOrder();
```

Returns default BitsOrder for fingerprints bit-vector fingerprints.

GetDefaultVectorStringFormat

```
$StringFormat = GetDefaultVectorStringFormat();
```

Returns default VectorStringFormat for fingerprints vector strings.

GetFingeprintsStringDelimiter

```
$Delimiter = GetFingeprintsStringDelimiter();
```

Returns string Delimiter used to generate fingerprints bit-vector and vector strings.

GenerateFingerprintsBitVectorString

```
$FPString = GenerateFingerprintsBitVectorString($FPBitVectorObject,
            [$BitStringFormat, $BitsOrder]);
```

Returns a FingerprintsString generated using *FingerprintsBitVectorObject* and optionally specified *BitStringFormat* and *BitsOrder* values.

Possible *BitStringFormat* values: *BinaryString, Binary, Bin, HexadecimalString, Hexadecimal, or Hex*. Default *BitStringFormat* value: *BinaryString*.

Possible *BitsOrder* values: *Ascending or Descending*. Default *BitsOrder* value: *Ascending*.

GenerateFingerprintsVectorString

```
$FPString = GenerateFingerprintsVectorString($FPVectorObject,
            [$VectorStringFormat]);
```

Returns a FingerprintsString generated using *FingerprintsVectorObject* and optionally specified *VectorStringFormat*.

Possible *VectorStringFormat* values: *IDsAndValuesString, IDsAndValues, IDsAndValuesPairsString, IDsAndValuesPairs, ValuesAndIDsString, ValuesAndIDs, ValuesAndIDsPairsString, ValuesAndIDsPairs, ValuesString, Values*.

Default *VectorStringFormat* value: for *NumericalValues FPVectorType - IDsAndValuesString*; for all other *FPVectorType*s - *ValuesString*.

GenerateFingerprintsString

```
$FPString = GenerateFingerprintsBitVectorString($FPBitVectorObject,
            [$BitStringFormat, $BitsOrder]);

$FPString = GenerateFingerprintsVectorString($FPVectorObject,
            [$VectorStringFormat]);
```

Returns a FingerprintsString generated using *FingerprintsBitVectorObject* or *FingerprintsVectorObject* and optionally specified parameters.

GetFingerprintsStringTypeAndDescription

```
($FPType, $FPDescription) = GetFingerprintsStringTypeAndDescription(
                            $FPString);
```

Returns FingerprintsStringType and *FingerprintsStringDescription* strings for FingerprintsString corresponding to FingerprintsBitVectorObject or FingerprintsVectorObject.

GetFingerprintsStringValues

```
@FPStringValues = GetFingerprintsStringValues($FPString);
```

Parses FingerprintsString corresponding to FingerprintsBitVectorObject or FingerprintsVectorObject and returns its individual component values as an array.

ParseFingerprintsBitVectorString

```
$FPBitVectorObject = ParseFingerprintsBitVectorString($FPBitVectorString,
                     [$ValidateValues]);
```

Returns FingerprintsBitVectorObject generated by parsing *FingerprintsBitVectorString* with optional validation of its component values.

ParseFingerprintsString

```
$FPBitVectorObject = ParseFingerprintsBitVectorString($FPBitVectorString,
                     [$ValidateValues]);

$FPVectorObject = ParseFingerprintsVectorString($FPVectorString,
                  [$ValidateValues]);
```

Returns FingerprintsBitVectorObject or ***FingerprintsVectorObject*** generated by parsing *FingerprintsBitVectorString* or *FingerprintsVectorString* with optional validation of its component values.

ParseFingerprintsVectorString

```
$FPVectorObject = ParseFingerprintsVectorString($FPVectorString,
                  [$ValidateValues]);
```

Returns FingerprintsVectorObject generated by parsing *FingerprintsVectorString* with optional validation of its component values.

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

BitVector.pm, FingerprintsBitVector.pm, FingerprintsVector.pm, Vector.pm

COPYRIGHT

Copyright (C) 2017 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.