NAME

MACCSKeys

SYNOPSIS

use Fingerprints::MACCSKeys;

use Fingerprints::MACCSKeys qw(:all);

DESCRIPTION

MACCSKeys [Ref 45-47] class provides the following methods:

new, GenerateFingerprints, GenerateMACCSKeys, GetDescription, SetSize, SetType, StringifyMACCSKeys

MACCSKeys is derived from Fingerprints class which in turn is derived from ObjectProperty base class that provides methods not explicitly defined in MACCSKeys, Fingerprints or ObjectProperty classes using Perl's AUTOLOAD functionality. These methods are generated on-the-fly for a specified object property:

```
Set<PropertyName>(<PropertyValue>);
$PropertyValue = Get<PropertyName>();
Delete<PropertyName>();
```

For each MACCS (Molecular ACCess System) keys definition, atoms are processed to determine their membership to the key and the appropriate molecular fingerprints strings are generated. An atom can belong to multiple MACCS keys.

For *MACCSKeyBits* value of Type option, a fingerprint bit-vector string containing zeros and ones is generated and for *MACCSKeyCount* value, a fingerprint vector string corresponding to number of MACCS keys [Ref 45-47] is generated.

MACCSKeyBits or MACCSKeyCount values for Type along with two possible 166 | 322 values of Size supports generation of four different types of MACCS keys fingerprint: MACCS166KeyBits, MACCS166KeyCount, MACCS322KeyBits, MACCS322KeyCount.

The current release of MayaChemTools generates the following types of MACCS keys fingerprints bit-vector and vector strings:

FingerprintsBitVector; MACCSKeyBits; 166; HexadecimalString; Ascending; 000 000000021210210e845f8d8c60b79dffbfffd1

METHODS

new

```
$NewMACCSKeys = new MACCSKeys(%NamesAndValues);
```

Using specified *MACCSKeys* property names and values hash, new method creates a new object and returns a reference to newly created PathLengthFingerprints object. By default, the following properties are initialized:

GetDescription

```
$Description = $MACCSKeys->GetDescription();
```

\$MACCSKeys->GenerateMACCSKeys();

Returns a string containing description of MACCS keys fingerprints.

 ${\tt GenerateMACCSKeys}\ or\ {\tt GenerateFingerprints}$

print "\$MACCSKeys\n";

```
$MACCSKeys = $MACCSKeys->GenerateMACCSKeys();
```

Generates MACCS keys fingerprints and returns MACCSKeys.

For *MACCSKeyBits* value of Type, a fingerprint bit-vector string containing zeros and ones is generated and for *MACCSKeyCount* value, a fingerprint vector string corresponding to number of MACCS keys is generated.

MACCSKeyBits or MACCSKeyCount values for Type option along with two possible 166 | 322 values of Size supports generation of four different types of MACCS keys fingerprint: MACCS166KeyBits, MACCS166KeyCount, MACCS322KeyBits, MACCS322KeyCount.

Definition of MACCS keys uses the following atom and bond symbols to define atom and bond environments:

```
Atom symbols for 166 keys [ Ref 47 ]:

A : Any valid periodic table element symbol
Q : Hetro atoms; any non-C or non-H atom
X : Halogens; F, Cl, Br, I
Z : Others; other than H, C, N, O, Si, P, S, F, Cl, Br, I
Atom symbols for 322 keys [ Ref 46 ]:

A : Any valid periodic table element symbol
Q : Hetro atoms; any non-C or non-H atom
X : Others; other than H, C, N, O, Si, P, S, F, Cl, Br, I
Z is neither defined nor used
```

```
Bond types:
    - : Single
    = : Double
    T : Triple
    # : Triple
       : Single or double query bond
    % : An aromatic query bond
   None: Any bond type; no explicit bond specified
    $ : Ring bond; $ before a bond type specifies ring bond
    ! : Chain or non-ring bond; ! before a bond type specifies chain bond
    \ensuremath{\text{@}} : A ring linkage and the number following it specifies the
         atoms position in the line, thus @1 means linked back to the first
         atom in the list.
   Aromatic: Kekule or Arom5
   Kekule: Bonds in 6-membered rings with alternate single/double bonds
            or perimeter bonds
    Arom5: Bonds in 5-membered rings with two double bonds and a hetro
            atom at the apex of the ring.
MACCS 166 keys [ Ref 45-47 ] are defined as follows:
   Key Description
    1 ISOTOPE
   2 103 < ATOMIC NO. < 256
   3 GROUP IVA, VA, VIA PERIODS 4-6 (Ge...)
    4 ACTINIDE
    5 GROUP IIIB, IVB (Sc...)
    6 LANTHANIDE
   7 GROUP VB, VIB, VIIB (V...)
   8 QAAA@1
   9 GROUP VIII (Fe...)
   10 GROUP IIA (ALKALINE EARTH)
   11 4M RING
   12 GROUP IB, IIB (Cu...)
   13 ON(C)C
   14 S-S
   15 OC(0)0
   16 OAA@1
   17 CTC
   18 GROUP IIIA (B...)
   19 7M RING
   20 SI
   21 C=C(Q)Q
   22 3M RING
    23 NC(0)0
   24 N-O
   25 NC(N)N
   26 C$=C($A)$A
   27 I
   28 QCH2Q
   29 P
   30 CQ(C)(C)A
   31 QX
   32 CSN
    33 NS
   34 CH2=A
   35 GROUP IA (ALKALI METAL)
   36 S HETEROCYCLE
    37 NC(O)N
   38 NC(C)N
   39 OS(O)O
```

40 S-O

- 41 CTN
- 42 F
- 43 QHAQH
- 44 OTHER
- 45 C=CN
- 46 BR
- 47 SAN
- 48 OQ(O)O
- 49 CHARGE
- 50 C=C(C)C
- 51 CSO
- 52 NN
- 53 QНАААОН
- 54 QHAAQH
- 55 OSO
- 56 ON(O)C
- 57 O HETEROCYCLE
- 58 QSQ
- 59 Snot%A%A
- 60 S=0
- 61 AS(A)A
- 62 A\$A!A\$A
- 63 N=O
- 64 A\$A!S
- 65 C%N
- 66 CC(C)(C)A
- 67 QS
- 68 QHQH (&...)
- 69 QQH
- 70 QNQ
- 71 NO
- 72 OAAO
- 73 S=A
- 74 CH3ACH3
- 75 A!N\$A
- 76 C=C(A)A
- 77 NAN
- 78 C=N
- 79 NAAN
- 80 NAAAN
- 81 SA(A)A
- 82 ACH2QH 83 QAAAA@1
- 84 NH2
- 85 CN(C)C
- 86 CH2QCH2
- 87 X!A\$A
- 88 S
- 89 OAAAO
- 90 QHAACH2A
- 91 QHAAACH2A
- 92 OC(N)C
- 93 QCH3
- 94 QN
- 95 NAAO
- 96 5M RING
- 97 NAAAO
- 98 QAAAAA@1
- 99 C=C
- 100 ACH2N
- 101 8M RING
- 102 QO
- 103 CL
- 104 QHACH2A
- 105 A\$A(\$A)\$A
- 106 QA(Q)Q
- 107 XA(A)A
- 108 CH3AAACH2A
- 109 ACH20

```
110 NCO
111 NACH2A
112 AA(A)(A)A
113 Onot%A%A
114 CH3CH2A
115 CH3ACH2A
116 CH3AACH2A
117 NAO
118 ACH2CH2A > 1
119 N=A
120 HETEROCYCLIC ATOM > 1 (&...)
121 N HETEROCYCLE
122 AN(A)A
123 OCO
124 QQ
125 AROMATIC RING > 1
126 A!O!A
127 A$A!O > 1 (&...)
128 ACH2AAACH2A
129 ACH2AACH2A
130 QQ > 1 (&...)
131 QH > 1
132 OACH2A
133 A$A!N
134 X (HALOGEN)
135 Nnot%A%A
136 O=A > 1
137 HETEROCYCLE
138 QCH2A > 1 (&...)
139 OH
140 0 > 3 (&...)
141 CH3 > 2 (&...)
142 N > 1
143 A$A!O
144 Anot%A%Anot%A
145 6M RING > 1
146 0 > 2
147 ACH2CH2A
148 AQ(A)A
149 CH3 > 1
150 A!A$A!A
151 NH
152 OC(C)C
153 QCH2A
154 C=O
155 A!CH2!A
156 NA(A)A
157 C-O
158 C-N
159 0 > 1
160 CH3
161 N
162 AROMATIC
163 6M RING
164 0
165 RING
166 FRAGMENTS
```

MACCS 322 keys set as defined in tables 1, 2 and 3 [Ref 46] include:

```
o 26 atom properties of type P, as listed in Table 1
o 32 one-atom environments, as listed in Table 3
o 264 atom-bond-atom combinations listed in Table 4
```

Total number of keys in three tables is: 322

Atom symbol, X, used for 322 keys [Ref 46] doesn't refer to Halogens as it does for 166 keys. In order to keep the definition of 322 keys consistent with the published definitions, the symbol X is used to imply "others" atoms, but it's internally mapped to symbol X as defined for 166 keys during the generation of key values.

Atom properties-based keys (26):

Key Description A(AAA) or AA(A)A - atom with at least three neighbors 1 Q - heteroatom 3 Anot%not-A - atom involved in one or more multiple bonds, not aromatic A(AAAA) or AA(A)(A)A - atom with at least four neighbors 4 5 A(QQ) or QA(Q) - atom with at least two heteroatom neighbors A(QQQ) or QA(Q)Q - atom with at least three heteroatom neighbors 6 QH - heteroatom with at least one hydrogen attached 8 CH2(AA) or ACH2A - carbon with at least two single bonds and at least two hydrogens attached 9 CH3(A) or ACH3 - carbon with at least one single bond and at least three hydrogens attached 10 Halogen 11 A(-A-A-A) or A-A(-A)-A - atom has at least three single bonds AAAAAA@1 > 2 - atom is in at least two different six-membered rings 12 13 A(\$A\$A\$A) or A\$A(\$A)\$A - atom has more than two ring bonds 14 A\$A!A\$A - atom is at a ring/chain boundary. When a comparison is done with another atom the path passes through the chain bond. Anot%A%Anot%A - atom is at an aromatic/nonaromatic boundary. When a 15 comparison is done with another atom the path passes through the aromatic bond. 16 A!A!A - atom with more than one chain bond 17 A!A\$A!A - atom is at a ring/chain boundary. When a comparison is done with another atom the path passes through the ring bond. 18 A%Anot%A%A - atom is at an aromatic/nonaromatic boundary. When a comparison is done with another atom the path passes through the nonaromatic bond. 19 HETEROCYCLE - atom is a heteroatom in a ring. 2.0 rare properties: atom with five or more neighbors, atom in four or more rings, or atom types other than H, C, N, O, S, F, Cl, Br, or I 21 rare properties: atom has a charge, is an isotope, has two or more multiple bonds, or has a triple bond. 22 N - nitrogen 23 S - sulfur 24 0 - oxygen 25 A(AA)A(A)A(AA) - atom has two neighbors, each with three or

Atomic environments properties-based keys (32):

more neighbors (including the central atom).

CHACH2 - atom has two hydrocarbon (CH2) neighbors

```
Description
Key
27
      C(CC)
28
      C(CCC)
29
      C(CN)
30
      C(CCN)
31
      C(NN)
32
      C(NNC)
33
      C(NNN)
34
      C(CO)
35
      C(CCO)
36
      C(NO)
37
      C(NCO)
38
      C(NNO)
39
      C(00)
40
      C(C00)
41
      C(NOO)
42
      C(000)
43
      Q(CC)
44
      O(CCC)
45
      Q(CN)
46
      O(CCN)
47
      Q(NN)
48
      Q(CNN)
49
      O(NNN)
50
      Q(CO)
51
      Q(CCO)
```

26

52 Q(NO) 53 Q(CNO) 54 Q(NNO) 55 Q(OO) 56 Q(COO) 57 Q(NOO) 58 Q(OOO)

Note: The first symbol is the central atom, with atoms bonded to the central atom listed in parentheses. Q is any non-C, non-H atom. If only two atoms are in parentheses, there is no implication concerning the other atoms bonded to the central atom.

Atom-Bond-Atom properties-based keys: (264)

```
Кеу
      Description
59
      C-C
60
      C-N
61
      C-0
62
      C-S
63
      C-Cl
      C-P
64
65
      C-F
66
      C-Br
67
      C-Si
68
      C-I
69
      C-X
70
      N-N
71
      N-O
72
      N-S
73
      N-Cl
74
      N-P
75
      N-F
76
      N-Br
77
      N-Si
78
      N-I
79
      N-X
80
      0-0
81
      0-S
82
      O-Cl
83
      0-P
84
      O-F
85
      0-Br
86
      O-Si
87
      O-I
88
      O-X
89
      S-S
90
      S-Cl
91
      S-P
92
      S-F
93
      S-Br
94
      S-Si
95
      S-I
96
      S-X
97
      Cl-Cl
98
      Cl-P
99
      Cl-F
100
      Cl-Br
      Cl-Si
101
102
      Cl-I
103
      Cl-X
104
      P-P
105
      P-F
106
      P-Br
107
      P-Si
108
      P-I
109
      P-X
110
      F-F
111
      F-Br
112
      F-Si
```

113

F-I

114 F-X115 Br-Br 116 Br-Si 117 Br-I 118 Br-X 119 Si-Si 120 Si-I Si-X 121 122 I-I 123 I-X 124 X-X 125 C=C 126 C=N 127 C=O 128 C=S 129 C=Cl 130 C=P 131 C=F 132 C=Br 133 C=Si 134 C=I 135 C=X 136 N=N 137 N=O 138 N=S 139 N=Cl 140 N=P 141 N=F 142 N=Br 143 N=Si 144 N=I145 N=X 146 0=0 147 0=S 148 O=Cl 149 O=P 150 O=F 151 0=Br 152 O=Si 153 O=I 154 O=X 155 S=S 156 S=Cl 157 S=P 158 S=F 159 S=Br 160 S=Si 161 S=I 162 S=X 163 Cl=Cl 164 Cl=P 165 Cl=F 166 Cl=Br Cl=Si 167 168 Cl=I 169 Cl=X170 P=P 171 P=F 172 P=Br 173 P=Si 174 P=I 175 P=X 176 F=F177 F=Br 178 F=Si 179 F=I180 F=X181 Br=Br

182

Br=Si

183 Br=I 184 Br=X 185 Si=Si 186 Si=I 187 Si=X 188 I = I189 I=X190 X=X 191 C#C 192 C#N 193 C#0 194 C#S 195 C#Cl 196 C#P 197 C#F 198 C#Br 199 C#Si 200 C#I 201 C#X 202 N#N 203 N#O 204 N#S 205 N#Cl 206 N#P 207 N#F 208 N#Br 209 N#Si 210 N#I 211 N#X 212 0#0 213 0#S 214 O#Cl 215 O#P 216 O#F 217 O#Br 218 O#Si 219 O#I 220 O#X 221 S#S 222 S#Cl 223 S#P 224 S#F 225 S#Br 226 S#Si 227 S#I 228 S#X 229 Cl#Cl 230 Cl#P 231 Cl#F 232 Cl#Br 233 Cl#Si 234 Cl#I 235 Cl#X 236 P#P 237 P#F 238 P#Br 239 P#Si 240 P#I 241 P#X 242 F#F243 F#Br 244 F#Si 245 F#I 246 F#X 247 Br#Br 248 Br#Si 249 Br#I 250 Br#X

251

Si#Si

252 Si#I 253 Si#X 254 I#I 255 I#X 256 X#X257 C\$C 258 C\$N 259 C\$0 260 C\$S 261 C\$Cl 262 C\$P 263 C\$F 264 C\$Br 265 C\$Si 266 C\$I 267 C\$X 268 N\$N 269 N\$O 270 N\$S 271 N\$Cl 272 N\$P 273 N\$F 274 N\$Br 275 N\$Si 276 N\$I 277 N\$X 278 0\$0 279 0\$S 280 O\$Cl 281 O\$P 282 O\$F 283 O\$Br 284 O\$Si 285 O\$I 286 O\$X 287 S\$S 288 S\$Cl 289 S\$P 290 S\$F 291 S\$Br 292 S\$Si 293 S\$I 294 S\$X 295 Cl\$Cl 296 Cl\$P 297 Cl\$F 298 Cl\$Br 299 Cl\$Si 300 Cl\$I 301 Cl\$X 302 P\$P 303 P\$F 304 P\$Br 305 P\$Si 306 P\$I 307 P\$X 308 F\$F 309 F\$Br 310 F\$Si 311 F\$I 312 F\$X 313 Br\$Br 314 Br\$Si 315 Br\$I 316 Br\$X 317 Si\$Si 318 Si\$I 319 Si\$X

320

I\$I

321 I\$X 322 X\$X

SetSize

\$MACCSKeys->SetSize(\$Size);

Sets size of MACCS keys and returns MACCSKeys. Possible values: 166 or 322.

SetType

\$MACCSKeys->SetType(\$Type);

Sets type of MACCS keys and returns MACCSKeys. Possible values: MACCSKeysBits or MACCSKeysCount.

StringifyMACCSKeys

```
$String = $MACCSKeys->StringifyMACCSKeys();
```

Returns a string containing information about MACCSKeys object.

AUTHOR

Manish Sud <msud@san.rr.com>

SEE ALSO

Fingerprints.pm, FingerprintsStringUtil.pm, AtomNeighborhoodsFingerprints.pm, AtomTypesFingerprints.pm, EStateIndiciesFingerprints.pm, ExtendedConnectivityFingerprints.pm, PathLengthFingerprints.pm, TopologicalAtomPairsFingerprints.pm, TopologicalAtomTripletsFingerprints.pm, TopologicalAtomTorsionsFingerprints.pm, TopologicalPharmacophoreAtomPairsFingerprints.pm, TopologicalPharmacophoreAtomPairsFingerprints.pm

COPYRIGHT

Copyright (C) 2020 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.