

```

1  '''
2      Matrix.py
3
4      Matrix class - Calculates the Minimum Edit Distance and generates the MED
5      Matrix, Pointer Matrix, and Edit String Alignment.
6  '''
7
8  from .Util import BTMATRIX_SYMBOLS, CellFormat
9  from .Cell import Cell
10 import os
11
12 class Matrix(object):
13     '''
14     Minimum Edit Distance Matrix Calculator
15     '''
16
17     def __init__(self, source_word, dest_word):
18         '''
19         Create a Minimum Edit Distance Matrix.
20             source_word - word to edit from
21             dest_word - word to edit to
22         '''
23         self._source_word = source_word
24         self._dest_word = dest_word
25         self._edit_string = str()
26
27         self._matrix_width = len(dest_word) + 1
28         self._matrix_height = len(source_word) + 1
29
30         # The matrix is stored as a list of lists.
31         # It is a list of rows, so the first coordinate
32         # is the row number and the second coordinate is
33         # the column number, resulting in the coordinate pair
34         # of (y, x). Noting this because it is counter to
35         # what is expected.
36         self._matrix = list()
37
38         # Now do the MED calculation and backtrace
39         self._doMEDCalculation()
40         self._generateEditString()
41
42
43     def _doMEDCalculation(self):
44         '''
45         Initialize the matrix and do the Minimum Edit Distance Calculation
46         For internal use only - Called by Matrix.__init__()
47         '''
48         # Initialize matrix with cells
49         for row in range(0, self._matrix_height):
50             self._matrix.append(list()) # Add a new row
51             for col in range(0, self._matrix_width):
52                 self._matrix[row].append(Cell())
53
54             # If the current cell is the origin or on the top or left
55             # boarder, initialize the value.
56             if row == 0: # initialize the first row
57                 self._matrix[row][col].distance = col
58             elif col == 0: # initialize the first column
59                 self._matrix[row][col].distance = row

```

```

60
61     # If not a boarder cell, calculate the cell value
62     else:
63         upvalue = self._matrix[row-1][col].distance + 1
64         leftvalue = self._matrix[row][col-1].distance + 1
65         diagvalue = self._matrix[row-1][col-1].distance
66
67         # Check to see if diagonal value is a substitution or not
68         # If it is a substitution, add 2 to diagonal value
69         if self._source_word[row-1] == self._dest_word[col-1]:
70             self._matrix[row][col].equivalent = True
71         else:
72             diagvalue += 2
73
74         # Calculate the minimum value and assign the distance to
75         # the cell
76         minvalue = min(upvalue, leftvalue, diagvalue)
77         self._matrix[row][col].distance = minvalue
78
79         # Now set the backtrace flags for the cell
80         if minvalue == upvalue:
81             self._matrix[row][col].up = True
82         if minvalue == leftvalue:
83             self._matrix[row][col].left = True
84         if minvalue == diagvalue:
85             self._matrix[row][col].diag = True
86
87
88     def _generateEditString(self):
89         '''
90         Preform the backtrace calculation and store the edit string
91         For internal use only - Called by Matrix.__init__()
92         '''
93         cell_row = self._matrix_height - 1
94         cell_col = self._matrix_width - 1
95         edit_string = str()
96         current_cell = self._matrix[cell_row][cell_col]
97
98         while (cell_row, cell_col) != (0, 0):
99             if current_cell.diag:
100                 if current_cell.equivalent:
101                     edit_string += " "
102                 else:
103                     edit_string += "s"
104                 cell_row -= 1
105                 cell_col -= 1
106             elif current_cell.up:
107                 edit_string += "d"
108                 cell_row -= 1
109             elif current_cell.left:
110                 edit_string += "i"
111                 cell_col -= 1
112             else:
113                 if cell_row == 0:
114                     edit_string += "i"
115                     cell_col -= 1
116                 elif cell_col == 0:
117                     edit_string += "d"
118                     cell_row -= 1

```

```

119         current_cell = self._matrix[cell_row][cell_col]
120
121         # Now reverse the string generated during the backtrace and store it
122         # String reversed using annoying python list extended slice syntax
123         self._edit_string = edit_string[::-1]
124
125     def _genFormattedMatrix(self, fun):
126         """
127         Generate a formatted matrix based upon the return of fun, which is a
128         lambda. For internal use only.
129         returns a string
130         """
131         # Note: As the formatted matrix looks like this:
132         #   #   #   B   R   I   E   F
133         #   #   -   -   -   -   -   -
134         #   D   -   -   -   -   -   -
135         #   R   -   -   -   -   -   -
136         #   I   -   -   -   -   -   -
137         #   V   -   -   -   -   -   -
138         #   E   -   -   -   -   -   -
139         #
140         # We have to play with off-by-one coordinates into the distance matrix.
141         # Also, as the matrix does not start printing the characters of the
142         # words until the 3rd row or column, word accesses are i-2
143         #
144         # Also, using lambdas! Learning Erlang was actually good for
145         # something!
146
147         return_string = str()
148         for row in range(0, self._matrix_height + 1):
149             for col in range(0, self._matrix_width + 1):
150                 # Print the # symbol to start the matrix
151                 if (row, col) == (0, 0) or (row, col) == (0, 1) or (row, col) == (1,
152 0):
153                     return_string += CellFormat('#')
154                     elif row == 0:
155                         return_string += CellFormat(self._dest_word[col - 2])
156                     elif col == 0:
157                         return_string += CellFormat(self._source_word[row - 2])
158                     else:
159                         return_string += CellFormat(fun(row-1, col-1))
160                 if row != self._matrix_height:
161                     return_string += os.linesep
162
163         return return_string
164
165     @property
166     def MED(self):
167         """
168         Returns the calculated Minimum Edit Distance
169         """
170         return self._matrix[self._matrix_height-1][self._matrix_width-1].distance
171
172     def getDistMatrixFormatted(self):
173         """
174         Returns the Distance Matrix as a formatted string for printing
175         """
176         return self._genFormattedMatrix(lambda y, x:

```

```

177         self._matrix[y][x].distance)
178
179
180     def getBTMatrixFormatted(self):
181         """
182         Returns the Backtrace Matrix as a formatted string for printing
183         """
184         return self._genFormattedMatrix(lambda y, x:
185             BTMATRIX_SYMBOLS[self._matrix[y][x].backtrace_flags])
186
187
188     def getStringAlignmentFormatted(self):
189         """
190         Returns the String alignment as a formatted string
191         """
192         source_word_counter = dest_word_counter = 0
193         aligned_source_word = str()
194         aligned_dest_word = str()
195         bar_string = "|" * len(self._edit_string)
196
197         for op in self._edit_string:
198             if op == "d":
199                 aligned_source_word += self._source_word[source_word_counter]
200                 source_word_counter += 1
201                 aligned_dest_word += "*"
202             elif op == "i":
203                 aligned_source_word += "*"
204                 aligned_dest_word += self._dest_word[dest_word_counter]
205                 dest_word_counter += 1
206             else:
207                 aligned_source_word += self._source_word[source_word_counter]
208                 source_word_counter += 1
209                 aligned_dest_word += self._dest_word[dest_word_counter]
210                 dest_word_counter += 1
211
212         return (aligned_source_word + os.linesep
213             + bar_string + os.linesep
214             + aligned_dest_word + os.linesep
215             + self._edit_string)

```