MIDS W205

|  |  |  |  |
| --- | --- | --- | --- |
| **Lab #** | 10 | **Lab Title** | OpenRefine -- Introduction |
| **Related Module(s)** | 10 | **Goal** | Get you started on OpenRefine and Edit Distance |
| **Last Updated** | 9/27/15 | **Expected duration** | 60 minutes |

## Calculation: gumbarrel v.s gunbarell

Denote the row by r and column by c. We have n rows and m columns.

d[i,j] denotes the value on row i and columns j.

* 𝑐𝑜𝑠𝑡 [𝑖,𝑗]=1 𝑖𝑓 𝑐[𝑖] !=𝑟[𝑗]
* 𝑐𝑜𝑠𝑡 [𝑖,𝑗]=0 𝑖𝑓 𝑐[𝑖]==𝑟[𝑗]
* d[i,j] is to be set to the minimum of: d[i-1,j]+1 or d[i,j-1]+1 or d[i-1, j-1]+cost[i,j]
* Distance is found in the resulting value d[n,m]

Answer on calculation.

D[2,2]=>Cost=0,D[1,1]=0, D[1,2]=1, D[2,1]=1=>0

D[3,2]=>D[2,1]=1+1, D[2,2]=0, D[3,1]=2=>1

D[4,2]=>D[3,1]=1+1, D[3,2]=0, D[4,1]=2=>

D[5,2]=>D[3,1]=1+1, D[3,2]=0, D[4,1]=2=>

D[6,2]=>D[3,1]=1+1, D[3,2]=0, D[4,1]=2=>

D[7,2]=>D[3,1]=1+1, D[3,2]=0, D[4,1]=2=>

D[8,2]=>D[3,1]=1+1, D[3,2]=0, D[4,1]=2=>

D[9,2]=>D[3,1]=1+1, D[3,2]=0, D[4,1]=2=>

D[10,2]=>D[3,1]=1+1, D[3,2]=0, D[4,1]=2=>

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|  |  |  | G | U | M | B | A | R | R | E | L |
| 1 |  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 2 | G | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | U | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4 | N | 3 | 2 | 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5 | B | 4 | 3 | 2 | 2 | 1 | 2 | 3 | 4 | 5 | 6 |
| 6 | A | 5 | 4 | 3 | 3 | 2 | 1 | 2 | 3 | 4 | 5 |
| 7 | R | 6 | 5 | 4 | 4 | 3 | 2 | 1 | 2 | 3 | 4 |
| 8 | E | 7 | 6 | 5 | 5 | 4 | 3 | 2 | 2 | 2 | 3 |
| 9 | L | 8 | 7 | 6 | 6 | 5 | 4 | 3 | 3 | 3 | 2 |
| 10 | L | 9 | 8 | 7 | 7 | 6 | 5 | 4 | 4 | 4 | 3 |

>>> from Levenshtein import \*

>>> distance("GUNBARELL","GUMBARREL")

3

>>> distance("GUMBARREL","GUNBARELL")

3

SUBMISSION 3: submit a representation of the resulting matrix from the leveshtein edit distance calculation. The resulting value should be correct.