

String Editing

M.B.Chandak

hodcs@rknec.edu

www.mbchandak.com

String Editing: Spell Checking

- Given two string: Source String and Destination String.
- The Source String will never modified.
- It is required to convert the destination string into source string using three possible operations: INSERT/DELETE/UPDATE.
- The cost of operation: INSERT=1, DELETE=1, UPDATE=2.
- The objective is to perform string editing in minimum cost.
- The process of translation/mapping is carried out phase wise.

String Editing: Cost of conversion

- Let the source string be $Y [y_1 \dots y_j]$
- Let the destination string be $X [x_1 \dots x_i]$
- Cost of transformation: depends on character at X_i and Y_j
- Insert/Delete/Update operation
- (i) If no string in Y and character present in X , to make $Y=X$: Perform?
- (ii) If string is in Y and no string is present in X , to make $Y=X$: Perform?
- (iii) If string contents present in Y and X , then cost of transformation depends upon:

String Editing: Cost of conversion: Formulation

$$\text{cost}(i, j) = \begin{cases} 0 & i = j = 0 \\ \text{cost}(i - 1, 0) + D(x_i) & j = 0, i > 0 \\ \text{cost}(0, j - 1) + I(y_j) & i = 0, j > 0 \\ \text{cost}'(i, j) & i > 0, j > 0 \end{cases}$$

where $\text{cost}'(i, j) = \min \left\{ \begin{array}{l} \text{cost}(i - 1, j) + D(x_i), \\ \text{cost}(i - 1, j - 1) + C(x_i, y_j), \\ \text{cost}(i, j - 1) + I(y_j) \end{array} \right\}$

String Editing: Cost matrix: Example 1

		0	1s	2s	3s	4s
			B	A	B	B
0d		0	1/i	2/l	3/l	4/i
1d	A	1/d	2/idu	1/u	2/l	3/i
2d	A	2/d	3/id	2/u	3/id	4/idu
3d	B	3/d	2/u	3/id	2/u	3/u
4d	A	4/d	3/d	2/u	3/id	4/idu
5d	B	5/d	4/u	3/d	2/u	3/u

A	2/IDU Assume "d" is used Delete character Cost=1 Move upwards Now insert character B Cost=1
A	Same character no operation/no cost use direction [U] COST=0
B	Same character no operation/no cost use direction [U] COST=0
A	Different characters [3id] Use "d". So delete character COST=1 [Move in upward row, as "d" is used]
B	Same character no operation and no cost. Use direction [u] [diagonal] COST=0

U	D
I	

String Editing: Cost matrix: Example 1 [option 2]

		0	1s	2s	3s	4s
			B	A	B	B
0d		0	1/i	2/l	3/l	4/i
1d	A	1/d	2/idu	1/u	2/l	3/i
2d	A	2/d	3/id	2/u	3/id	4/idu
3d	B	3/d	2/u	3/id	2/u	3/u
4d	A	4/d	3/d	2/u	3/id	4/idu
5d	B	5/d	4/u	3/d	2/u	3/u

A	2/D Assume "d" is used Delete character Cost=1 Move upwards 1/D Delete character Cost=1
B	Same character no operation/no cost use direction [U] COST=0
B	Same character no operation/no cost use direction [U] COST=0
A	Different characters [3id] Use "l". So INSERT character COST=1 [Move in SIDE, as "l" is used]
B	Same character no operation and no cost. Use direction [u] [diagonal] COST=0

U	D[i-1]
I[j-1]	

String Editing: Cost of conversion

- Let the source string be $Y [y_1 \dots y_j] = b a b b$
- Let the destination string be $X [x_1 \dots x_i] = a a b a b$

$i \downarrow j \rightarrow$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	1	2	3
2	2	3	2	3	4
3	3	2	3	2	3
4	4	3	2	3	4
5	5	4	3	2	3

Home work

- Source String: a b a a b a b
- Destination string: a b a a a b a
- It is required to convert the destination string to source string

- Source String: E D I T I N G
- Destination string: E D G I N G / E A T I N G / D E N T I N G
- :It is required to convert the destination string to source string

- ***Note take source string in Columns of matrix***