Levenshtein Edit Distance

Given two strings u and v over an alphabet Σ, the Levenstein distance from u to v (or from v to u) is the number of edit steps needed to change u to v, where an edit step is one of the following:

1. Delete a symbol.
2. Insert a symbol.
3. Replace a symbol.

For examle, the Levenshtein distance from WARM to BEAR is 3, since we can change WARM to BEAR with three edit steps:

WARM
1. delete M: WAR 2. replace W with B: BAR replace 3. insert E: BEAR

Levenstein distance is computed using dynamic programming. Let n be the length of u and m the length of v. Oet $u[i]$ be the prefix of u of length i and let $v[j]$ be the prefix of v of length j, for $0 \leq i \leq n$ and $0 \leq j \leq m$.

Subproblem(i,j) is defined to be the computation of the Levenstein distance from $u[i]$ to $v[j]$, which we call $L[i,j]$. here are (n+1)(m+1) subproblems. The final answer is $L[n,m]$.

The program is as follows. Let u_i, v_i be the i^{th} symbol of u and the j^{th} symbol of v, respectively. The program is as follows:

For all i let $L[i,0] = i$
For all j let $L[0,j] = j$
For all $1 \leq i \leq n$
    For all $1 \leq j \leq n$
        For all $1 \leq j \leq n$
            If(u[i] = v[j])
                L[i,j] = min{L[i-1,j]+1,L[i,j-1]+1,L[i-1,j-1]}
            else
                L[i,j] = min{L[i-1,j]+1,L[i,j-1]+1,L[i-1,j-1]+1}

Example

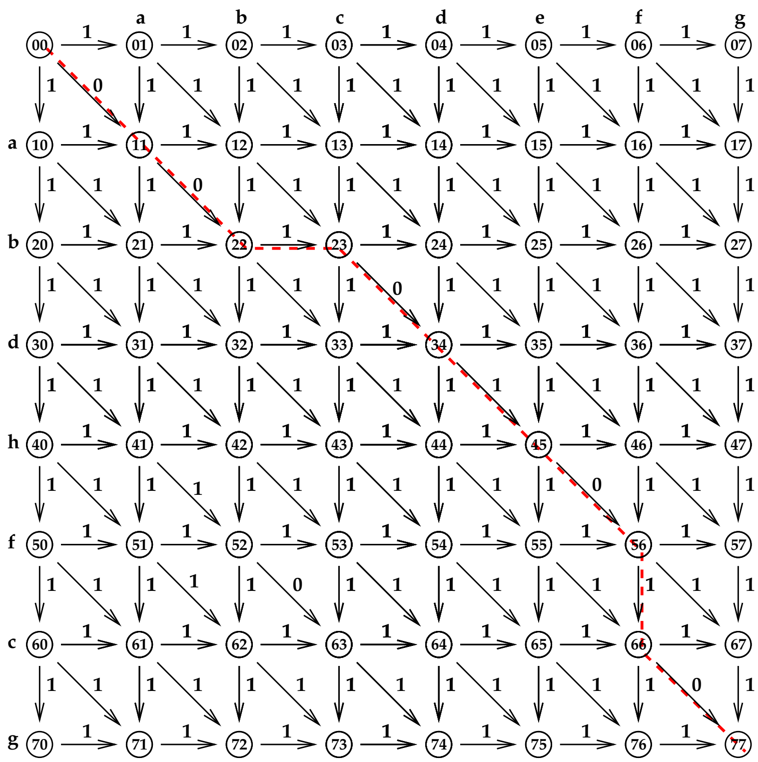
The following matrix shows the values of L.

			B	E	A	R
		0	1	2	3	4
	0	0	1	2	3	4
W	1	1	1	2	3	4
A	2	2	2	2	2	3
R	3	3	3	3	3	2
M	4	4	4	4	4	3

The Levenshein distance is $L[4,4] = 3$

Another Example

Compute the Levenshtein edit distance from abdhfbg to abcdefg. The Levenstein edit problem reduces to the single source shortest path problem. Finding the edit distance between strings of length m and n reduces to finding the shortest distance between two vertices in a weighted directed graph with $(m+1)(n+1)$ vertices. We show that graph below, as well as the shortest path. The source is labeled 00, and the shortest path from 00 to ij is the value of $L[i,j]$ in the pseudocode, which is the entry in the i^{th} row, j^{th} column of the matrix.



		a	b	c	d	e	f	g
	0	1	2	3	4	5	6	7
a	1	0	1	2	3	4	5	6
b	2	1	0	1	2	3	4	5
d	3	2	1	1	1	2	3	4
h	4	3	2	2	2	2	3	4
f	5	4	3	3	3	3	2	3
c	6	5	4	3	4	4	3	3
g	7	6	5	5	5	5	4	3

The Levenstein distance is $L[7,7] = 3$. The steps are:

1. insert c between b and d,

2. change h to e,

3. delete the c after f.