

Given the two sequences of letters:

s1: boeken

s2: books

1. Create a distance matrix

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

2. Provide two possible alignments and corresponding transcripts with the minimum edit cost.

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

Transcript: [“match” (b), “match” (o),
“substitution” (e into o), “match” (k),
“deletion” (e), “substitution” (n into s)]

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

Transcript: [“match” (b), “insert” (o),
“match” (o), “deletion” (e), “match” (k),
“deletion” (e), “substitution” (n into s)]

Alignment:

b o e k e n
 | | | | | |
 b o o k * s
 s d s

Alignment:

b * o e k e n
 | | | | | |
 b o o * k * s
 i d d s

3. Are there still more alignments with minimal edit cost possible? How can you see?

Yes. During the creation of the minimal-cost path (starting from the top right) we can make several choices in diversity of points along the matrix. For example, in the first cell (corresponding to “n” and “s”) we can chose either first insert and then a delete, or first delete and then insert, or a direct substitution. All three operations have the same cost, which, as we can see, it is 5.

Another critical point along the matrix is the cell that represents the “e” and the “o”. Whatever path we decide to follow in the first cell (“n” and “s”) we end up in this new cell in the middle, where we can make three new choices. Again, it does not matter which path we follow, all three have the minimal edit cost.

Each choice we make result in a different path to continue along the matrix, therefore, it multiplies the number of alignments that we can achieve with the minimal cost of operations.

Just considering these 2 cells (3 possible paths each) we have already 9 possible paths ($3 \times 3 = 9$).

However, 9 are not all the possible paths we can follow, there are more. For instance, notice that depending on the path we decide to take we will need to make more decisions!

4. If there are more alignments with minimal edit cost, how many of those are there in total?

There are 20 possible alignments with the minimal edit cost. To understand it better, I represented each path in the matrix dividing them in three big groups:

Yellow: possible m.e.c. alignments if we decide to start with substitution: substitute “n” for “s” (consider that we start at the end of the word – top right cell)

c	6	7	5	7	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
b	1	2	0	1	1	2	2	3	3	4	4
#	0	1	2	3	4	5					
#	b	o	o	k	s						

n	6	7	5	7	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
b	1	2	0	1	1	2	2	3	3	4	4
#	0	1	2	3	4	5					
#	b	o	o	k	s						

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

Green: possible m.e.c. alignments if we decide to start with insertion:

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	4	2	3	3
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

Blue: possible m.e.c. alignments if we decide to start with deletion. Notice that if we start with deletion, the next cell we encounter demands to make three more choices. Find the possibilities organized with the same logic as before: first substitution, second insertion, and last, deletion.

- Substitution

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
	#	b	o	o	k	s					

- Insertion

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
#	b	o	o	k	s						

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
#	b	o	o	k	s						

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
#	b	o	o	k	s						

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
#	b	o	o	k	s						

- deletion

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
#	b	o	o	k	s						

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
#	b	o	o	k	s						

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
#	b	o	o	k	s						

n	6	7	5	7	4	5	5	6	4	5	5
		7	5	6	4	5	5	6	4	5	5
e	5	6	4	5	3	4	4	5	3	4	4
		6	4	5	3	4	4	5	3	4	4
k	4	5	3	4	2	3	3	4	2	3	3
		5	3	4	2	3	3	2	4	5	5
e	3	4	2	3	1	2	2	3	3	4	4
		4	2	3	1	2	2	3	3	4	4
o	2	3	1	2	0	1	1	2	2	3	3
		3	1	0	2	1	3	4	4	5	5
b	1	2	0	1	1	2	2	3	3	4	4
		0	2	3	3	4	4	5	5	6	6
#	0	1	2	3	4	5					
#	b	o	o	k	s						