

РЕДАКЦИОННОЕ РАССТОЯНИЕ

(расстояние Левенштейна) между строками s и t — это минимальное количество операций замены (Change), вставки (Insert) или удаления (Delete) символа, которые надо совершить, чтобы перевести s в t .

Задача о нахождении редакционного расстояния, как и любая задача динамического программирования, сводится к нахождению стоимости оптимального пути в ориентированном ациклическом графе подзадач.

Пусть $D_{i,j}$ — редакционное расстояние между префиксами $s[0..i]$ и $t[0..j]$ длин i и j . Ясно, что $D_{i,0} = i$, $D_{0,j} = j$. Другие значения определяются рекуррентным соотношением

$$D_{i,j} = \begin{cases} D_{i-1,j-1}, & \text{если } s[i] = t[j]; \\ 1 + \min\{D_{i-1,j}, D_{i,j-1}, D_{i-1,j-1}\} & \text{иначе.} \end{cases}$$

Пример: $s = \text{ЭССЕ}$, $t = \text{СЕТ}$, редакционное расстояние равно 3.

		Э	С	С	Е
	0	1	2	3	4
С	1	1	1	2	3
Е	2	2	2	2	2
Т	3	3	3	3	3

Редакционное предписание — строка из символов **M** (Match), **C** (Change), **I** (Insert) и **D** (Delete), описывающая последовательность действий, необходимых для получения одной строки из другой, и содержащая минимальное количество букв **C**, **I**, **D**.

Редакционное предписание можно получить, выполнив обратный ход по вычисленной таблице расстояний.

Начиная с правого нижнего угла таблицы, рисуем стрелки во все соседние клетки, соответствующие подзадачам меньшего размера и меньшим редакционным расстояниям, или

по диагонали из клетки, соответствующей одинаковым буквам («бесплатный переход»).

		Э	С	С	Е
	0	1	2	3	4
С	1	1	1	2	3
Е	2	2	2	2	2
Т	3	3	3	3	3

		Э	С	С	Е
	0	1	2	3	4
С	1	1	1	2	3
Е	2	2	2	2	2
Т	3	3	3	3	3

Движение вверх означает вставку буквы в исходное слово (**I**), движение влево — удаление буквы (**D**), движение по диагонали — это либо замена буквы (**C**), либо «бесплатный переход» в случае совпадения букв (**M**).

		Э	С	С	Е
	0	1	2	3	4
С	1	1	1	2	3
Е	2	2	2	2	2
Т	3	3	3	3	3

		Э	С	С	Е
	0	1	2	3	4
С	1	1	1	2	3
Е	2	2	2	2	2
Т	3	3	3	3	3

Количество возможных редакционных предписаний равно количеству путей в графе подзадач, имеющих минимальную стоимость.

		Э	С	С	Е
	0	1	2	3	4
С	1	1	1	2	3
Е	2	2	2	2	2
Т	3	3	3	3	3

РЕДАКЦИОННЫЕ ПРЕДПИСАНИЯ	
D D M M I	
D M D M I	
D M C C	

В а р и а н т ы з а д а ч и:

1. Разные цены замены, вставки, удаления букв, а также «бесплатного перехода» при их совпадении. Например, задача о расстоянии Хэмминга и задача о длине наибольшей общей подпоследовательности — это частные случаи задачи о редакционном расстоянии.

2. Расстояние Дамерау—Левенштейна: добавляется операция перестановки соседних букв.

		П	И	Р	А	М	И	Д	А
	0	1	2	3	4	5	6	7	8
П	1	0	1	2	3	4	5	6	7
Р	2	1	1	1	2	3	4	5	6
И	3	2	1	1	2	2	3	4	5
З	4	3	2	2	2	3	3	4	5
М	5	4	3	3	3	3	4	4	5
А	6	5	4	4	3	4	4	5	5

M D M C I M D D M

		П	И	Р	А	М	И	Д	А
	0	1	2	3	4	5	6	7	8
П	1	0	1	2	3	4	5	6	7
Р	2	1	1	1	2	3	4	5	6
И	3	2	1	1	2	2	3	4	5
З	4	3	2	2	2	3	3	4	5
М	5	4	3	3	3	3	4	4	5
А	6	5	4	4	3	4	4	5	5

Match ↓
Delete ↓
Delete ↓
Match ↓
Insert ↓
Change ↓
Match ↓
Delete ↓
Match ↓

П - Р И З М - - А

		П	О	Л	Е
	0	1	2	3	4
К	1	1	2	3	4
О	2	2	1	2	3
Р	3	3	2	2	3
О	4	4	3	3	3
В	5	5	4	4	4
А	6	6	5	5	5

C M I C C I

		П	О	Л	Е
	0	1	2	3	4
К	1	1	2	3	4
О	2	2	1	2	3
Р	3	3	2	2	3
О	4	4	3	3	3
В	5	5	4	4	4
А	6	6	5	5	5

Insert ↓
Change ↓
Change ↓
Insert ↓
Match ↓
Change ↓

К О Р О В А