

Редакционное расстояние (Расстояние Левенштейна)

Задача о выравнивании

$S_1 = \text{SUNNY}$

3

$S_2 = \text{SNOWY}$

$S_1 = \text{РОЖДЬ}$

3

$S_2 = \text{ПРОЖЬ}$

SUNNY

3

SNOWY

РОЖДЬ

2

ПРОЖЬ

Задача о редакционном расстоянии:

Вход S_1, S_2 .

Три операции:

1. замена

2. удаление

3. вставка

a
b
a
-
-
b

Вопрос: минимальное кол-во операций

$S_1 \rightsquigarrow S_2$ и последовательность операций

$E(S_1, S_2) = ?$

$S_1 = \text{POLYNOMIAL}$

$S_2 = \text{EXPONENTIAL}$

Решение динамическим программированием.

$$1. E[i, j] = E(S_1[1:i], S_2[1:j])$$

$$2. E[0, j] = j \quad E[i, 0] = i$$

$$E[i, j] = \min \left\{ E[i-1, j-1] + [S_1[i] \neq S_2[j]], \right.$$

POLY
EXP-

$$E[i-1, j] + 1, \text{ // удаление}$$

$$E[i, j-1] + 1 \text{ // вставка}$$

3. Порядок: по столбцам / по строкам

	EXPONENTIAL											
	0	1	2	3	4	5	6	7	8	9	10	11
P	1	1	2	2	3							
O	2	2	2	3	2							
L	3	3	3	3	3							
Y	4	4	4	4	4							
N	5	5	5	5	5							
O	6	6	6	6	5							
M	7	7	7	7	6							
I	8	8	8	8	7							
A	9	9	9	9	8							
L	10	10	10	10	9							

"Рашируемые" стрелочки
в обратном порядке =>
рекурсивная последовательность

- ↘ замена
- ↓ удаление
- вставка

Стоимость операции —
веса на ребрах

Время: $O(|S_1| \cdot |S_2|)$ Память: $O(|S_1| \cdot |S_2|)$

Задача:

Если хранить только последний столбец =>
Память: $O(\min\{|S_1|, |S_2|\})$, но нельзя
восстановить рекурсивно последовательность

Алгоритм Хиршберга

Идея: meet in the middle

1. $E(S_1, S_2) = E(\text{rev}(S_1), \text{rev}(S_2))$

2. $S_1 = S_{11} \circ S_{12}$

∃ такое разбиение $S_2 = S_{21} \circ S_{22}$

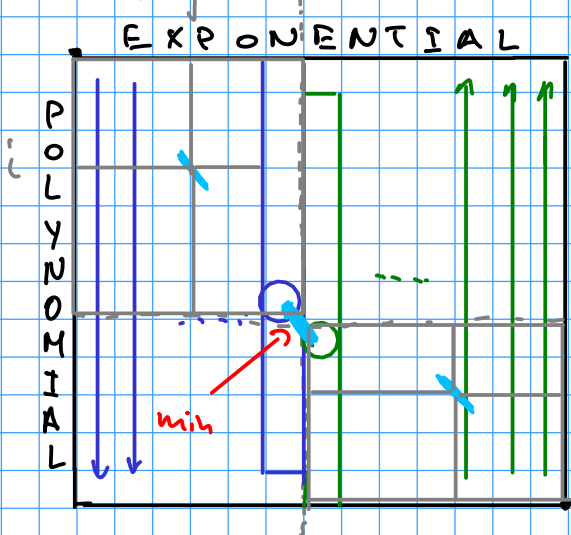
$E(S_1, S_2) = E(S_{11}, S_{21}) + E(S_{12}, S_{22})$

	S_{11}	S_{12}
S_1	POLYNOMIAL	
S_2	EXPONENTIAL	
	S_{21}	S_{22}

"Доказательство"

DP	OX	AB
DP	OX	B

1. Расстояние не увеличится при "правильном" разрезании, т.к. стандарт сохраняется
2. Расстояние не уменьшится, т.к. в противном случае мы могли бы уменьшить исходное выражение. \square



Итерация № 1.

1. Разбиваем S_2 пополам

$$S_{21} = \text{EXPON}$$

$$S_{22} = \text{ENTIAL}$$

2. Решаем две задачи:

$$E(S_1, S_{21})$$

$$E(\text{rev}(S_1), \text{rev}(S_{22}))$$

Результат:

i :

$$E(S_1[i:i], S_{21}) + E(S_1[i+1:n], S_{22}) \rightarrow \min$$

3. Находим оптимальное разделение строки

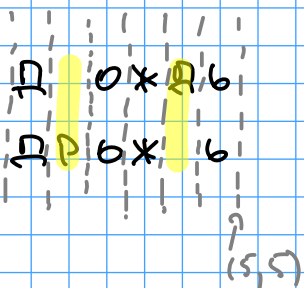
S_1 , пог. коэф. \min

цели в коэф. единиц

Средняя итерация:

рекурсивно вызывает две (S_{11}, S_{21}) и (S_{12}, S_{22})

Выход: список пар индексов, задающих оптимальное выражение



Анализ: Размер $O(\min\{S_1, S_2\})$

$$\text{Время: } |S_1| \cdot |S_2| \cdot \left(1 + \frac{1}{2} + \frac{1}{4} + \dots\right) = O(|S_1| \cdot |S_2|)$$

