Количество выравниваний.

Лазарева София

8 февраля 2023 г.

Условие:

Выведите рекуррентную формулу количества всех возможных выравниваний последовательностей длины n и m пользуясь разбиением всех выравниваний на непересекающиеся блоки. (1.5 балл)

Получите точную формулу, основываясь на начальные условия и рекуррентную формулу. (1.5 балл)

Воспользуйтесь приближением Стирлинга чтобы получить приближенную формулу количества выравниваний. (1)

Решение:

1. Применим динамику. Для того, чтобы посчитать число всех возможных выравниваний строк $|a| = n, |b| = m$ воспользуемся динамикой по подстрокам.

Базой динамики будет число выравниваний строк $a_{1...i}$ и b_1 для всех $i = 1...n$ равное единице. Аналогично число выравниваний для a_1 и $b_{1...j}$ для всех $j = 1...m$.

Предположим, что на момент подсчёта всех выравниваний строк $a_{1...i}$ и $b_{1...j}$, мы уже знаем количество выравниваний строк $a_{1...i-1}, b_{1...j}$, $a_{1...i-1}, b_{1...j-1}$, $a_{1...i}, b_{1...j-1}$. Тогда для того, чтобы посчитать число выравниваний строк $a_{1...i}$ и $b_{1...j}$ необходимо просуммировать выравнивания их подстрок, т.е. `count[i][j] = count[i-1][j] + count[i-1][j-1] + count[i][j-1]`.

Число выравниваний исходных строк будет лежать в элементе `count[n][m]`.

2. Обозначим $M = \min(n, m)$, а $N = \max(n, m)$. Тогда если записать данную рекуррентную формулу через биномиальные коэффициенты, то получим следующее соотношение:

$$\texttt{count[i][j]} = \sum_{i=0}^{M} C_M^i \cdot C_{N+M-i}^M = \sum_{i=0}^{M} \frac{(N+M-i)!}{i! \cdot (M-i)! \cdot (N-i)!}$$

3. Воспользуемся приближением Стирлинга для вычисления факториалов: $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$:

$$\texttt{count[i][j]} = \sum_{i=0}^{M} \frac{\sqrt{2\pi(M+N-i)}(M+N-i)^{M+N-i}}{\sqrt{2\pi(M-i)}\sqrt{2\pi(N-i)}\sqrt{2\pi i}(M-i)^{M-i}(N-i)^{N-i}(i)^i}$$