

Семинар по префикс-функции

Сапожников Денис

Задача 1. Дана строка s и текст t . Требуется найти количество вхождений строки s в текст t за $O(|s| + |t|)$.

Решение. Запишем строку $s\#t$ и вычислим на ней префикс-функцию, где $\#$ — символ-разделитель, который не встречается в s и t . Если на какой-то позиции i : $\pi_i = |s|$, то это означает, что текущий суффикс строки t равен строке s , а это то, что мы и хотели найти. Три человека трудилось в поте лица над этим алгоритмом и теперь у него есть название: **алгоритм Кнута-Морриса-Пратта**.

Задача 2. Пусть у вас есть функция `isSubstring`, которая проверяет, является ли одна строка подстрокой другой строки. Вам нужно определить, является ли строка s_1 циклическим сдвигом строки s_2 , используя только один вызов функции `isSubstring`.

Задача 3. Дана строка s . Для каждого префикса s необходимо найти количество его вхождений за.

1. $O(|s|^2)$
2. $O(|s|)$

Задача 4. Дана строка s . Необходимо найти наименьший период строки¹ за линейное время.

Задача 5. Придумайте линейный по сложности алгоритм для вычисления количества различных циклических сдвигов строки s .

Задача 6. Даны строки p и s . s состоит только из маленьких букв английского алфавита, в то время как p может содержать ещё и символы '*'. Требуется определить, можно ли подставить вместо символов '*' какие-то строки (возможно, различные), так чтобы строка p стала подстрокой s .

1. $O(|s|^2 + |p| \cdot |s|)$
2. $O(|s| + |p|)$

Например, если $p = \text{«*hello**!»}$, а $s = \text{«hello world!!!!!»}$, то первую и третью звездочку можно превратить в пустую строку, вторую — в « world!!!!!» и тогда строки совпадут.

Есть ещё одна интересная теорема, которая очень часто применяется в теории строк, но не в этой лекции.

Задача 7 (Fine and Wilf's theorem). Если у строки w есть периоды p и q , где $|w| \geq p + q - \gcd(p, q)$, то $\gcd(p, q)$ также является периодом этой строки.

Зачем она нужна? С помощью неё можно умудриться считать префикс-функцию в онлайн с добавлением и удалением символов за $O(\log n)$.

¹Периодом строки s называется такое наименьшее число T , что $\forall i : s_i = s_{i+T}$