

# Алгоритм Ахо-Корасик. Суффиксные структуры

Минский ШАД. Осень

27 марта 2015 г.

## 1 Примечания

Напоминаю, что суффиксный массив человечество умеет строить за линейное время от длины строки, чем мы и будем пользоваться. Тем не менее в виду сложности данного алгоритма в практической части разрешается использовать алгоритм построения за  $\mathcal{O}(n \log n)$ .

## 2 Тематические задачи

1. [1 ½ балла] Маленькому Сэмюэлю на день рождения подарили набор  $A$  из  $N$  кодовых слов, причём суммарная длина всех слов равна  $L$ . Он хочет проверить, правда ли, что набор задаёт однозначно декодируемый код, т.е. из того, что  $a_1 a_2 \dots a_n = b_1 b_2 \dots b_m$ , где  $\forall i : a_i \in A, b_i \in A$  верно, что  $n = m$  и  $\forall i : a_i = b_i$ .

У малыша не так много времени, поэтому алгоритм должен иметь сложность  $\mathcal{O}(LN)$ .

2. [1 балл] Маленький Джордж загадал бинарную строку из  $N$  бит. Он посчитал по ней суффиксный массив и отдал вам его. Сможете ли вы отгадать строку, которую загадал малыш? Кстати говоря, он мог ошибиться и предоставить вам массив, которому не соответствует ни одна бинарная строка, тогда надо указать ему на ошибку. Тем не менее, малыш умён не по годам, поэтому если вы назовёте строку, которая порождает такой же массив, как и загаданная, то он по доброте душевной сочтёт загадку разгаданной.

Если вы будете решать загадку слишком долго, малыш решит, что вы очень скучный человек и пойдёт к студентам киевского филиала, так что решите задачу за  $\mathcal{O}(N)$ .

3. [½ балла] Малыши Майкл и Ричард играют в следующую игру. Майкл придумал строку из  $S$  символов и дал Ричарду примерно  $\mathcal{O}(|S| \log |S|)$  чтобы вдоволь её изучить. После этого, он задаёт вопросы вида, правда ли, что в позиции  $i$  строки  $S$  начинается tandemный повтор длины  $k$ . Другими словами, правда ли, что  $S[i \dots i + k - 1] = S[i + k \dots i + 2k - 1]$ . Ричард не хочет опозориться перед Майклом (в будущем им, возможно, придётся вместе писать статьи), поэтому он просит Вас помочь отвечать на каждый такой вопрос за  $\mathcal{O}(1)$ .

## 3 Задачи на повторение

4. [1 балл] Малышу Палу подарили массив из  $nm + 1$  различных чисел. Он знает, что не так интересен массив чисел, как возрастающие (ну или хотя бы убывающие последовательности). Однако малыш, имени которого, к сожалению наш герой не разобрал, сказал ему, что тот никогда не найдёт в своём массиве возрастающей последовательности длины  $n + 1$  и уж тем более убывающей последовательности длины  $m + 1$ . Малыш Иоганн тут же успокоил Пала, сказав что это не так. Станьте номером один для Пала и докажите, что слова Иоганна не пустой звук, т.е. всегда найдётся хотя бы одна из указанных последовательностей.

5. [1 балл] Малыш Чарльз подарил малышам Энтони и Ричарду массив из  $n = 2^k$  различных целых чисел. А как известно, чем меньше числа, тем лучше. Энтони очень заботится о малыше Ричарде, поэтому уступает ему минимум из подаренного массива. Тем не менее, ему тоже очень хочется узнать, какой же подарок получит он. А значит, он хочет определить 2-ю порядковую статистику за не более, чем  $n + k - 2$  сравнения (больше ждать он уж не в силах). Помогите ему составить алгоритм, который найдёт 2-ю порядковую статистику не более чем за приведённое количество сравнений.

## 4 Практические задачи

Ссылка на констест: <https://contest.yandex.ru/contest/1080/problems/>

6. [1 балл] Реализуйте решение задачи 3.
7. [1 балл] Дана строка длины  $n$  и её суффиксный массив. Требуется найти количество различных строк в массиве, максимум в массиве LCP и количество различных значений функции  $LCP(i, j)$  за  $\mathcal{O}(n)$
8. [1 балл] Малыш Абрахам получил в подарок от малыша Якоба в подарок строку. Его друг малыш Терри тут же решил загадать Абрахаму загадку. А именно его интересует, как сильно похоже строка, начиная с позиции  $i$  на какую-нибудь подстроку, начинающуюся в более ранней позиции. А именно, хочется найти  $f(i)$  для всех возможных  $i$ , если  $f(i) = \max\{k : \exists j < i, s[j, j+k] = s[i, i+k]\}$ .
9. [2 балла] Дано  $n$  строк общей длины  $T$ . Также дано  $m$  запросов вида в скольки различных данных строках есть подстрока  $Q_i$ . Строки  $Q_i$  образуют беспрефиксный код. Общий размер запросов не превышает  $S$ . Необходимо ответить на все запросы за время  $\mathcal{O}(S + T)$

Задание	1	2	3	4	5	6	7	8	9	Сумма
Баллы	1½	1	½	1	1	1	1	1	2	10