

Суффиксные деревья

Необходимо реализовать алгоритм Укконена построения суффиксного дерева за линейное время. Построив такое дерево для некоторых из входных строк, необходимо воспользоваться полученным суффиксным деревом для решения своего варианта задания.

Алфавит строк: строчные буквы латинского алфавита (т.е., от а до z).

Упрощенный вариант

Реализовать поиск подстрок в тексте с использованием суффиксного дерева. Суффиксное дерево можно построить за $O(n^2)$ наивным методом. Задание, алфавит, входные/выходные данные и пример такие же, как в первом варианте.

Вариант 1: Поиск в известном тексте неизвестных заранее образцов

Найти в заранее известном тексте поступающие на вход образцы.

Входные данные: текст располагается на первой строке, затем, до конца файла, следуют строки с образцами.

Выходные данные: для каждого образца, найденного в тексте, нужно распечатать строчку, начинающуюся с последовательного номера этого образца и двоеточия, за которым, через запятую, нужно перечислить номера позиций, где встречается образец в порядке возрастания.

Входной файл	Выходной файл
abcdabc	1: 1
abcd	2: 2
bcd	3: 2, 6
bc	

Вариант 2: Поиск с использованием суффиксного массива

Найти в заранее известном тексте поступающие на вход образцы с использованием суффиксного массива. Формат входных и выходных данных, а так же примеры аналогичны указанным во первом задании.

Вариант 3: Поиск образца с использованием статистики

совпадений

Найти образец в тексте используя статистику совпадений.

Входные данные: на первой строке располагается образец, на второй — текст.

Выходные данные: последовательность строк содержащих в себе номера позиций, начиная с которых встретился образец. Строки должны быть отсортированы в порядке возрастания номеров.

Входной файл	Выходной файл
aba	2
qababaz	4

Вариант 4: Линеаризация циклической строки

Линеаризовать циклическую строку, то есть найти минимальный в лексикографическом смысле разрез циклической строки.

Входные данные: некий разрез циклической строки.

Выходные данные: минимальный в лексикографическом смысле разрез.

Входной файл	Выходной файл
xabcd	abcdx

Вариант 5: Поиск наибольшей общей подстроки

Найти самую длинную общую подстроку двух строк.

Входные данные: две строки.

Выходные данные: на первой строке нужно распечатать длину максимальной общей подстроки, затем перечислить все возможные варианты общих подстрок этой длины в порядке лексикографического возрастания без повторов.

Входной файл	Выходной файл
xabay	3
xabcbay	bay
	xab