МФТИ
Алгоритмы и структуры данных, осень 2022
Семинар №06. Суффиксный автомат (2)

Всюду в этом листке считаем, что алфавит — константного размера.

1. Найдите k-ю (без учёта числа вхождений) подстроку s в лексикографическом порядке за $O(|s|)$.
2. Для каждой подстроки строки s найдите её первое и последнее вхождение в s, а также количество её вхождений.
3. За $O(|s| + |t|)$ найдите наибольшую общую подстроку строк s и t.
4. Найдите наибольшую общую подстроку строк $s_1, \ldots, s_k$ за время $O(k \cdot \sum\limits_{i=1}^{k} |s_i|)$.
5. Докажите, что дерево суффиксных ссылок в суффиксном автомате для строки s совпадает с суффиксным деревом для строки s^R.
6. Дана строка $s = s_1 \ldots s_n$ и набор строк-запросов $t_1, \ldots, t_q$. Для каждого i сообщите количество таких j, что $s_j \ldots s_{j+|t_i|-1}$ равна какому-либо циклическому сдвигу t_i. Сдвиги могут быть разными для разных j. Асимптотика: $O\left(n + \sum\limits_{i=1}^{q} |t_i|\right)$.
7. Пусть s — строка. Скажем, что строка y предваряет строку x, если непосредственно перед любым вхождением x в s встречается вхождение y. Найдите число пар (x, y), таких что y предваряет x. Как быть с парами (x, y), в которых, наоборот, y продолжает x (то есть непосредственно за каждым вхождением x следует вхождение y)?
8. Покажите, как по суффиксному автомату построить суффиксный массив за линейное время.

1. Каждая подстрока соответствует пути в автомате. Из каждой вершины найдите количество путей.
2. Первое вхождение строки v соответствует длиннейшему пути из v в терминальную вершину. Последнее — кратчайшему. Количество вхождений можно насчитать динамикой на развёрнутом графе.
3. Для каждого префикса строки t найдите его самый длинный суффикс, являющийся подстрокой s. Для этого можно отрезать символы от начала.
4. Постройте суффиксный автомат для строки $s_1 + d_1 + s_2 + d_2 + \ldots + s_k + d_k$, где все d_i — попарно различные разделители. Затем для каждого состояния и каждого разделителя требуется определить существование некоторого пути из этого состояния, содержащего только этот разделитель.
5. Вспомните критерий того, что строка является `longest` в своём классе эквивалентности.
6. Постройте суффиксный автомат по строке s. Для каждого t_i пройдитесь строкой $t_i + t_i$ по этому автомату (обрубайте первые символы, чтобы была строка длины $|t_i|$).
7. Остановимся на случае предваряющих y. Тогда x и yx имеют одинаковое множество вхождений. Значит, нужно сфокусироваться на каждом состоянии суффиксного автомата s отдельно. Чтобы решить задачу для продолжающих y, достаточно развернуть строку s.
8. Сожмите проходные (вершины со входящей и исходящей степенью 1) нетерминальные вершины, запустите `dfs` без сохранения пометок `used`. Иными словами, пройдите в этом автомате все пути. Время работы будет пропорционально числу суффиксов и числу различных подстрок u, из которых возникает ветвление, то есть ua и ub есть подстроки s. Иными словами, u^R — `longest` в своём классе для строки s^R.