

КМП-алгоритм. Z-алгоритм

Минский ШАД. Осень

2 марта 2015 г.

1 Обозначения

Бордером строки будем называть такой её префикс, который совпадает с суффиксом. Например, у строки «abacaba» бордерами являются «а», «aba» и «abacaba». **Собственным бордером** будем называть бордер, меньший по длине, чем сама строка.

π -**последовательностью** строки S будем называть числовой вектор π размера $|S|$, такой что π_i равен длине самого длинного собственного бордера первых $i + 1$ символов строки (в ноль-индексации). Например для «abacaba» $\pi = \{0, 0, 1, 0, 1, 2, 3\}$.

z -**последовательностью** строки S будем называть числовой вектор z размера $|S|$, такой что z_i равен длине максимальной (по длине) подстроки, которая начинается в позиции i и совпадает с некоторым собственным суффиксом строки. Например для «abacaba» $z = \{0, 0, 1, 0, 3, 2, 1\}$.

2 Тематические задачи

1. [$\frac{1}{2}$ балла] Дана числовая последовательность. Проверить, правда ли, что существует строка, π -последовательность которой совпадает с данной последовательностью.
2. [$\frac{1}{2}$ балла] Дана π -последовательность строки. Предложить алгоритм поиска любой строки, порождающей данную последовательность и доказать корректность (вообще всегда надо доказывать, но тут я подчеркну).
3. [$\frac{1}{2}$ балла] Дана (изначально пустая) строка. Каждый ход к ней дописывается один символ в начало или в конец. После каждого хода за $\mathcal{O}(|S|)$ необходимо говорить, сколько различных подстрок существует в S . Предполагается решение с помощью КМП или z -алгоритма.
4. [$\frac{1}{2}$ балла] Предложить алгоритм поиска для строки второго по длине собственного бордера. Сложность должна быть $\mathcal{O}(|S|)$.
5. [$\frac{1}{2}$ балла] Предложить алгоритм вычисления количества различных бордеров у строки S за время $\mathcal{O}(|S|)$.
6. [$\frac{1}{2}$ балла] Для каждой позиции строки S вычислить значение a_i — длину максимальной подстроки, которая начинается в i и совпадает с некоторым суффиксом строки S . Решение должно иметь сложность $\mathcal{O}(n)$.
7. [1 балл] Дана строка S , пусть $|S| = n$. Затем следует запросы вида (i, j) , на каждый из которых надо ответить длину j -го бордера (считая бордеры упорядоченными по длине) у строки, которая является i -префиксом строки S , т.е. подстрока $S[0 \dots i - 1]$. На каждый запрос следует отвечать не медленней, чем за $\mathcal{O}(\log n)$. Разрешается сделать препроцесс за $\mathcal{O}(n \log n)$.
8. [1 балл] Рассмотрим следующую игру для двух игроков. Первый игрок загадывает строку S и сообщает второму её длину n . Также у первого игрока есть изначально пустая строка T . Затем игроки ходят по очереди начиная с первого. На своём ходу первый игрок добавляет в конец строки

T любую букву (строка T также неизвестна второму игроку). Второй игрок имеет право задать первому не более пяти вопросов вида «правда ли что такой-то символ строки S (T) совпадает с таким-то символом строки S (T)». Т.е. можно сравнивать любые позиции в одной и той же строке, либо в разных строках — единственное ограничение, что нельзя задавать более пяти вопросов за ход. В конце своего хода второй игрок обязан сказать, сколько на данный момент подстрок строки T совпадают со строкой S . Ваша задача разработать стратегию игры для второго игрока, чтоб его ответ всегда был правильным, вне зависимости от игры первого игрока, либо доказать, что такой стратегии не существует.

9. [1 1/2 балла] Дана π -последовательность строки. Посчитать количество строк, которые порождают данную последовательность над алфавитом размера C . Сложность алгоритма должна составлять $\mathcal{O}(|\pi| \log |\pi|)$. Можно считать, что искомое количество помещается в машинное слово (тем не менее, это не значит, что от него должна зависеть сложность).
10. [1 1/2 балла] Задана строка,жатая RLE-алгоритмом, т.е. последовательностью пар (c_i, l_i) — символ и количество повторений соответственно. Например строка «aaabbbaeee» будет закодирована такой последовательностью: $\{(a, 3), (b, 2), (a, 1), (e, 3)\}$. Таких пар — N . Также задано M вопросов (заранее), каково значение префикс-функции данной строки в позиции i . Предложить алгоритм ответа на эти запросы за время $\mathcal{O}(T \log^2 T)$, где $T = \max N, M$.
11. [1 балл] Доказать или опровергнуть следующие утверждения (π — π -последовательность, z — z -последовательность):
 - (a) $\sum \pi_i = \sum z_i$ для любой строки
 - (b) $\sum \pi_i < \sum z_i$ для любой строки
 - (c) $\sum \pi_i > \sum z_i$ для любой строки
 - (d) $\sum \pi_i \leq \sum z_i$ для любой строки
 - (e) $\sum \pi_i \geq \sum z_i$ для любой строки
 - (f) Существует строка, что $\pi_i > z_i$ для любого $i > 0$
 - (g) Существует строка, что $z_i > \pi_i$ для любого $i > 0$

3 Задачи на повторение

12. [1/2 балла] На прямой своими координатами задано n точек. В этих точках расположены гвоздики. Два гвоздика, находящихся в позициях x_i и x_j можно соединить ниткой длиной $|x_i - x_j|$ сажень. Необходимо натянуть нитки между гвоздями таким образом, чтоб к каждому гвоздю была присоединена как минимум одна нитка, а суммарная длина нитей была минимальна. Сложность алгоритма должна составлять $\mathcal{O}(n \log n)$.
13. [1/2 балла] Дано N пар натуральных чисел (a_i, b_i) . Необходимо посчитать количество различных пар натуральных чисел (A, B) , таких что, $\exists i : A \leq a_i \wedge B \leq b_i$ за время $\mathcal{O}(N \log N)$.

4 Практические задачи

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

14. [1 балл] Реализовать решение задач 1 и 2
15. [1 балл] По π -последовательности строки найти её z -последовательность (используй 2, Люк).

| | | | | | | | | | | | | | | | | |
|---------|-----|-----|-----|-----|-----|-----|---|---|-------|-------|----|-----|-----|----|----|-------|
| Задание | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | Сумма |
| Баллы | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1 | 1 | 1 1/2 | 1 1/2 | 1 | 1/2 | 1/2 | 1 | 1 | 12 |