

Первый курс, весенний семестр
Практика по алгоритмам #9
Строки: префикс-функция, Z, хеши

Contents

1	Новые задачи	2
2	Домашнее задание	3
2.1	Обязательная часть	3
2.2	Дополнительная часть	3

1 Новые задачи

1. gcd – тоже период!

Пусть у строки s есть периоды $a, b \leq \frac{|s|}{2}$. Докажите, что $\gcd(a, b)$ – тоже период.

2. В поисках периода.

- a) Найти кратчайший период строки тремя способами: КМП, Z-функция, хеши.
- b) Найти все периоды строки.

3. Подсчёт различных подстрок.

- a) Найти число различных подстрок строки. $\mathcal{O}(n^2)$. Два способа: Z-функция, хеши.
- b) Найти подстроку данной строки, встречающуюся максимальное число раз.

4. Позиция строки в суффиксном массиве.

Найти позицию строки в ее суффиксном массиве. Два способа: Z-функция, хеши.

5. k -й суффикс.

Найти k -й в лексикографическом порядке суффикс строки.

- a) $\mathcal{O}(n \log^2 n)$.
- b) $\mathcal{O}(n \log n)$.

6. Суффиксный массив и стандартные сортировки.

Построение суффиксного массива хешами за $\mathcal{O}(n \log^2 n)$:
что оптимальнее использовать `sort` или `stable_sort`?

7. Поиск с одной ошибкой.

Научиться искать образец в строке, если допустимо различие в один символ между образцом и найденной подстрокой.

8. Поиск с перестановками символов и алфавита.

Найти образец в строке, если допустимо:

- a) в образце применять к алфавиту перестановку.
- b) в образце переставлять символы.
- c) в образце переставлять и алфавит, и символы.

9. Восстановление строки по P и Z функциям.

За $\mathcal{O}(n)$ восстановить строку, если дана ее

- a) Z-функция.
- b) префикс-функция.

10. Наибольшая дважды подстрока.

Найти наибольшую по длине строку, которая дважды без перекрытий встречается в заданной строке. $\mathcal{O}(n \log n)$.

11. Палиндромы.

- a) Найти количество подпалиндромов строки. $\mathcal{O}(n \log n)$.
- b) Найти максимальный подпалиндром строки. $\mathcal{O}(n)$.

12. (*) Префиксы представимые в $\alpha\beta$ -виде.

Для каждого префикса строки проверить, представим ли он в виде $\alpha\beta\alpha\beta\dots\alpha$, где α и β – произвольные, возможно пустые, строки, строка β повторяется ровно k раз.

13. (*) Minimal cyclic shift.

Найти минимальный циклический сдвиг строки за $\mathcal{O}(n)$.

2 Домашнее задание

2.1 Обязательная часть

1. **(3) Тандемный повтор 1.**

Тандемным повтором называется строка вида $\alpha\alpha$. Найдите за $\mathcal{O}(n^2)$ самый длинный тандемный повтор. Нужно представить три решения, используя (а) хеши, (б) Z-функция, (с) предподсчитанный lcp. За каждое решение вы получите по баллу

2. **(3) Общий подпалиндром.**

Нужно за $\mathcal{O}(n \log n)$ найти максимальный общий подпалиндром.

3. **(3) Ретрострока.**

Для каждого префикса строки найти количество его префиксов равных его суффиксу. $\mathcal{O}(n)$.

4. **(3*) LZSS.**

Алгоритм кодирования LZSS. Дана строка s . Выписываем её слева направо. Пусть уже выписан префикс $[0, i)$. Можно или, потратив 1 доллар, записать в код строки s_i и выписать i -й символ, или, потратив 5 долларов, записать в код строки (j, len) и выписать сразу len символов. Здесь $j < i$, а $s[j : j+len) = s[i : i+len)$. Ваша задача – за $\mathcal{O}(n^2)$ выписать всю строку за минимальную стоимость. Дополнительный балл можно получить, решив задачу с $\mathcal{O}(n)$ памяти.

5. **(3) $Z \rightarrow \text{КМП}$**

Преобразовать Z-функция в префикс-функцию без промежуточного восстановления строки.

6. **(3) Поиск с ошибкой в алфавите.**

Найти подстроку в тексте. При сравнении строк можно делать циклический сдвиг алфавита в одной из них. $\mathcal{O}(n\Sigma)$. Здесь Σ – размер алфавита.

7. **(3) Поиск с двумя ошибками.**

Найти подстроку в тексте. При сравнении строк, если несовпадений было не более двух, строки считаются равными. $\mathcal{O}(n)$.

8. **(3) Поиск с k ошибками.**

Найти подстроку в тексте. При сравнении строк, если несовпадений было не более k , строки считаются равными. $\mathcal{O}(nk \log n)$.

2.2 Дополнительная часть

1. **(5) Обезьянка за клавиатурой.**

За одну секунду в конец изначально пустого текста дописывается случайная буква (равномерное распределение). Какое матожидание времени T , когда первый раз s станет подстрокой выписанного текста?

2. **(6) Антихеш тест.**

Даны целые числа p и m . Построить две разных строки, у которых (p, m) полиномиальный хеш совпадёт. $h(s_0, s_1, \dots, s_n) = (\sum s_i p^i) \bmod m$.

а) **(2)** $\mathcal{O}(m^{1/2})$.

б) **(2)** $\mathcal{O}(m^{1/3})$.

с) **(2)** $m \leq 10^{36}$.

3. **(4) Тандемный повтор 2.**

Решите задачу про тандемный повтор за $\mathcal{O}(n \log n)$ методом разделяй и властвуй!

4. **(4) Покрывание строки.**

Говорят, что строка α покрывает строку s , если каждый символ s покрыт хотя бы одним вхождением α . Дана s , найти минимальную по длине α . $\mathcal{O}(n \log n)$.