

# Суффиксный массив

Гусев Илья

Московский физико-технический институт

Москва, 2017

- 1 Суффиксный массив
  - Определение
  - Пример
  - Алгоритм построения за  $O(n \cdot \log(n))$
  - Алгоритм построения за  $O(n * \log(n))$

# Суффиксный массив - определение

Дана строка  $s$  длины  $n$ .

$i$ -ый суффикс -  $s_i = s[i \dots n - 1]$ .

Суффиксный массив:

$$suf = \{i \in [0 \dots n - 1] : \forall j \neq i \quad suf.index[i] < suf.index[j] \rightarrow s_i <_{lex} s_j\}$$

То есть, это перестановка индексов суффиксов  $s$ , которая задаёт порядок суффиксов в порядке лексикографической сортировки.

# Суффиксный массив - определение

Пример:

$s = abacaba$

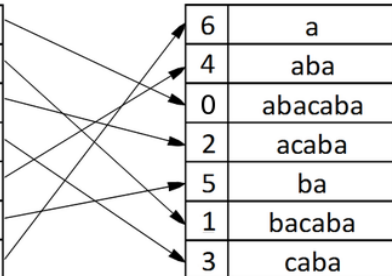
$suf = [6, 4, 0, 2, 5, 1, 3]$ .

Все суффиксы строки

0	abacaba
1	bacaba
2	acaba
3	caba
4	aba
5	ba
6	a

Лексикографический порядок

6	a
4	aba
0	abacaba
2	acaba
5	ba
1	bacaba
3	caba



# Алгоритм построения за $O(n \cdot \log(n))$

Особенности:

- 1 Дополняем до циклических сдвигов
- 2  $s$  - класс эквивалентности по  $2^{k-1}$  символам
- 3 Сортируем первые  $2^k$  символов на основе первых отсортированных  $2^{k-1}$  и природы суффиксов -  $O(\log(n))$  шагов
- 4 Сортировка пар цифр за  $O(n)$
- 5 Всего  $O(n * \log(n))$

# Алгоритм построения за $O(n * \log(n))$

Пошаговый пример - 0

$p$								
0	a	b	a	c	a	b	a	#
1	b	a	c	a	b	a	#	a
2	a	c	a	b	a	#	a	b
3	c	a	b	a	#	a	b	a
4	a	b	a	#	a	b	a	c
5	b	a	#	a	b	a	c	a
6	a	#	a	b	a	c	a	b
7	#	a	b	a	c	a	b	a

# Алгоритм построения за $O(n \cdot \log(n))$

## Пошаговый пример - 1

$p$	$c$								
7	0	#	a	b	a	c	a	b	a
0	1	a	b	a	c	a	b	a	#
2	1	a	c	a	b	a	#	a	b
4	1	a	b	a	#	a	b	a	c
6	1	a	#	a	b	a	c	a	b
1	2	b	a	c	a	b	a	#	a
5	2	b	a	#	a	b	a	c	a
3	3	c	a	b	a	#	a	b	a

# Алгоритм построения за $O(n \cdot \log(n))$

## Пошаговый пример - 1

$$k = 0$$

Для строки  $i$ :

$$class1_i = c[i]$$

$$class2_i = c[j : p_j = p_i + 2^k = p_i + 1]$$

$p$	$c$									class1	class2
7	0	#	a	b	a	c	a	b	a	0	1
0	1	a	b	a	c	a	b	a	#	1	2
2	1	a	c	a	b	a	#	a	b	1	3
4	1	a	b	a	#	a	b	a	c	1	2
6	1	a	#	a	b	a	c	a	b	1	0
1	2	b	a	c	a	b	a	#	a	2	1
5	2	b	a	#	a	b	a	c	a	2	1
3	3	c	a	b	a	#	a	b	a	3	1



# Алгоритм построения за $O(n \cdot \log(n))$

## Пошаговый пример - 2

Сортируем по паре (class1, class2)

$p$	$c$									class1	class2
7	0	#	a	b	a	c	a	b	a	0	1
6	1	a	#	a	b	a	c	a	b	1	0
0	1	a	b	a	c	a	b	a	#	1	2
4	1	a	b	a	#	a	b	a	c	1	2
2	1	a	c	a	b	a	#	a	b	1	3
1	2	b	a	c	a	b	a	#	a	2	1
5	2	b	a	#	a	b	a	c	a	2	1
3	3	c	a	b	a	#	a	b	a	3	1

# Алгоритм построения за $O(n \cdot \log(n))$

## Пошаговый пример - 2

$$k = 1$$

Для строки  $i$ :

$$class1_i = c[i]$$

$$class2_i = c[j : p_j = p_i + 2^k = p_i + 2]$$

$p$	$c$									class1	class2
7	0	#	a	b	a	c	a	b	a	0	4
6	1	a	#	a	b	a	c	a	b	1	2
0	2	a	b	a	c	a	b	a	#	2	3
4	2	a	b	a	#	a	b	a	c	2	1
2	3	a	c	a	b	a	#	a	b	3	2
1	4	b	a	c	a	b	a	#	a	4	5
5	4	b	a	#	a	b	a	c	a	4	0
3	5	c	a	b	a	#	a	b	a	5	4

# Алгоритм построения за $O(n \cdot \log(n))$

## Пошаговый пример - 3

$$k = 2$$

Для строки  $i$ :

$$class1_i = c[i]$$

$$class2_i = c[j : p_j = p_i + 2^k = p_i + 4]$$

$p$	$c$									class1	class2
7	0	#	a	b	a	c	a	b	a	0	5
6	1	a	#	a	b	a	c	a	b	1	3
4	2	a	b	a	#	a	b	a	c	2	3
0	3	a	b	a	c	a	b	a	#	3	2
2	4	a	c	a	b	a	#	a	b	4	1
5	5	b	a	#	a	b	a	c	a	5	6
1	6	b	a	c	a	b	a	#	a	6	5
3	7	c	a	b	a	#	a	b	a	7	0

# Алгоритм построения за $O(n \cdot \log(n))$

## Пошаговый пример - 3

Сортируем по паре (class1, class2)

$p$	$c$									class1	class2
7	0	#	a	b	a	c	a	b	a	0	5
6	1	a	#	a	b	a	c	a	b	1	3
4	2	a	b	a	#	a	b	a	c	2	3
0	3	a	b	a	c	a	b	a	#	3	2
2	4	a	c	a	b	a	#	a	b	4	1
5	5	b	a	#	a	b	a	c	a	5	6
1	6	b	a	c	a	b	a	#	a	6	5
3	7	c	a	b	a	#	a	b	a	7	0

# Полезные ссылки I



Е-maxx: суффиксный массив

[http://e-maxx.ru/algo/suffix\\_array](http://e-maxx.ru/algo/suffix_array)



Викиконспекты: суффиксный массив

[https://neerc.ifmo.ru/wiki/index.php?title=Суффиксный\\_массив](https://neerc.ifmo.ru/wiki/index.php?title=Суффиксный_массив)