

Суфф. массив (задачи)  
( 06.10.17 )

# 1. Поиск подстроки (напоминание)

Наивное решение с использованием суфф. массива

Асимптотика -  $O(|p| * \log|s|)$

mlr - оптимизация

# 1. Поиск подстроки. LCP (longest common prefix)

<b>lcp[i]</b>		<b>0</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>
<b>Suff[i]</b>	7	6	0	3	1	4	2	5
0	\$	a	a	a	a	a	b	c
1		\$	a	a	b	c	a	a
2			b	c	a	a	a	\$
3			a	a	a	\$	c	
4			a	\$	c		a	
5			c		a		\$	
6			a		\$			
7			\$					

# 1. Поиск подстроки. LCP (свойства)

$LCP(s, t)$  - длина наибольшего общего префикса строк  $s$  и  $t$ .

$$LCP(S_{Suf[x]}, S_{Suf[z]}) = \min_{i=x+1 \dots z} lcp[i]$$

# 1. Поиск подстроки. LCP (longest common prefix)

1. Построение суфф. массива
2. Построение массива LCP  
(алгоритм Касаи  $O(|s|)$ , лекция)
3. Бинарный поиск по суфф.массиву  
с использованием LCP

## 1. Поиск подстроки. LCP (бин. поиск)

Хотим найти диапазон индексов  $[Ans\_left, Ans\_right]$  в суфф. массиве, для которых соответствующие суффиксы содержат искомый *pattern* в качестве префикса

Разберем для  $Ans\_left$ .

# 1. Поиск подстроки. LCP (бин. поиск)

Исходные границы поиска  $l = 0$ ,  $r = |s| - 1$ .

Найдем LCP для *pattern* и соответствующих суффиксов за  $O(|pattern|)$ . Обозначим их:  $lcp\_l$ ,  $lcp\_r$

$$m = (l + r) / 2$$

Найдем (RMQ)  $m\_l = \text{LCP}(\text{suff}[l], \text{suff}[m])$ ,  
 $m\_r = \text{LCP}(\text{suff}[r], \text{suff}[m])$

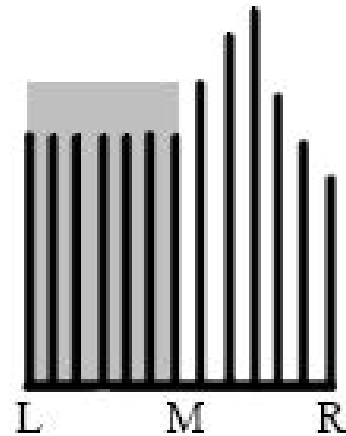
# 1. Поиск подстроки. LCP (бин. поиск, 1 случай)

без о.о.  $lcp_l \geq lcp_r$

1.  $m_l > lcp_l$ .

Сл-но продолжаем поиск справа.

$lcp_l$  остается без изменений



(1)

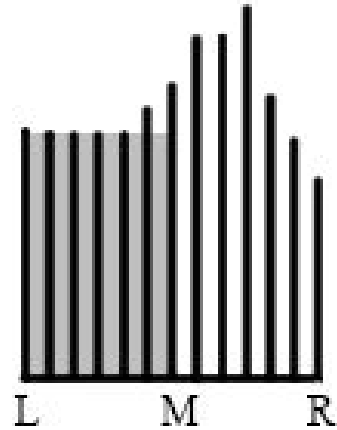


1. Поиск подстроки. LCP (бин. поиск, 2 случай)

2.  $m_l == lcp_l$ .

Сравниваем  $suff[m]$  с  $pattern$ ,  
начиная с  $lcp_l$ -ого символа.

В зависимости от результата  
сравнения выбираем нужную  
половину (соответствующее LCP  
для  $pattern$  пересчитано)

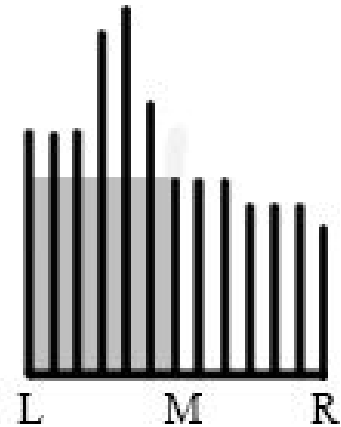


(2)

# 1. Поиск подстроки. LCP (бин. поиск, 3 случай)

3.  $m_l < lcp_l$ .

Продолжаем поиск слева.

$$lcp\_r = m\_l.$$


(3)

# 1. Поиск подстроки. Итог

Асимптотика:

Препроцессинг: суфф. массив + LCP

Запрос:  $O(\log|s| + |pattern|)$

[https://neerc.ifmo.ru/wiki/index.php?title=Алгоритм\\_поиска\\_подстроки\\_в\\_строке\\_с\\_помощью\\_суффиксного\\_массива](https://neerc.ifmo.ru/wiki/index.php?title=Алгоритм_поиска_подстроки_в_строке_с_помощью_суффиксного_массива)

## 2. Нахождение числа уникальных подстрок

1. Построение суфф. массива
2. Построение массива LCP
3. Подсчет результата

## 2. Нахождение числа уникальных подстрок

Проходимся по суффиксному массиву.

Для  $i$ -ого суффикса  $s_i$  добавляем  $|s_i| - \text{lcp}[i]$  новых уникальных подстрок.

Асимптотика -  $O(|s|)$  + суфф. массив