

Основа алгоритма КМП – Построение **префикс-функции**.

Префикс - подстрока, начинающаяся с **первого индекса строки**.

Суффикс – группа последовательных символов строки поиска, у которой **зафиксирован последний символ**.

Значением префикс-функции для каждого Суффикса является **максимальное количество совпадающих символов с начала префикса и конца суффикса**.

Рассмотрим строку:

a	a	b	a	a	b	a	a	a	a	b	a	a	b	a	a	a	b
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

— порядковый номер K символа в строке, начиная с 0

0	1	0	1	2	3	4	5	2	2	3	4	5	6	7	8	9	3
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

— массив M префиксов

(значений префикс-функции)

По правилам применения КМП-поиска значения префикс-функции показывают величину сдвига для образца при несовпадении.

Возьмем символ с номером **6** (это **a**) и для K от 0 до 5 рассмотрим **строки-префиксы** (подстрока, начинающаяся с первого индекса строки) и **суффиксы** (подстрока, последний символ которой в строке в позиции 6 (это символ **a**) длины $K+1$).

$K+1$	префикс	суффикс
1	a	a
2	aa	ba
3	aab	aba
4	aaba	aaba
5	aabaa	baaba
6	aabaab	abaaba

- префикс — первый символ строки, а суффикс - символ a

- самое длинное совпадение; здесь и ниже суффикс и префикс начали перекрываться (как фрагменты строки поиска)

Замечание. Нумерация символов строки, начинающаяся с 0, ориентирована на C^{++}

Очевидно, что при таком методе поиска образец (подстрока поиска) **обязательно** должен быть помещен в начало строки (текста).

Поиск очередного (кроме начального) вхождения образца можно осуществить следующим образом:

1. В строке поиска, начиная с символа, непосредственно следующего за образцом, ищется символ, совпадающий с конечным символом образца.
2. Если значение префикс-функции для этого элемента строки равно длине образца, фиксируется очередное вхождение образца в строку.
3. Поиск производится до окончания строки поиска.

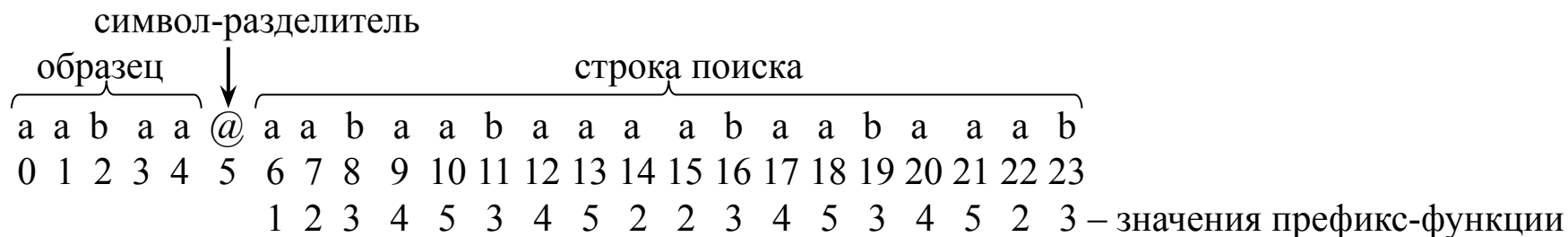
Замечание. Очевидно, что если значение префикс-функции для элемента строки, совпадающего с конечным символом образца, **больше** размера образца, совпадение не фиксируется.

Можно предложить другой подход, при котором префикс-функция строится **не для образца**, а для **строки поиска**.

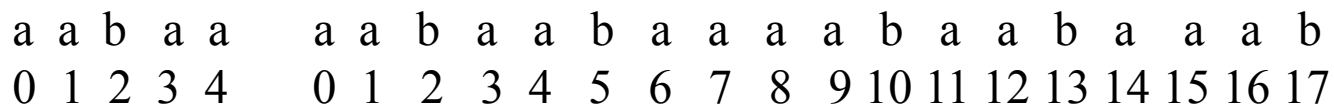
В этом случае, для поиска вхождений образца в исходную строку не нужно «двигать» образец по строке, а просто просмотреть элементы префикс-функции.

Очевидно, что каждое вхождение образца в строку фиксируется, когда значение элемента префикс-функции совпадает с длиной образца.

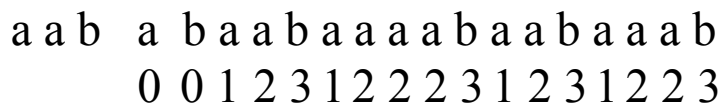
Пример



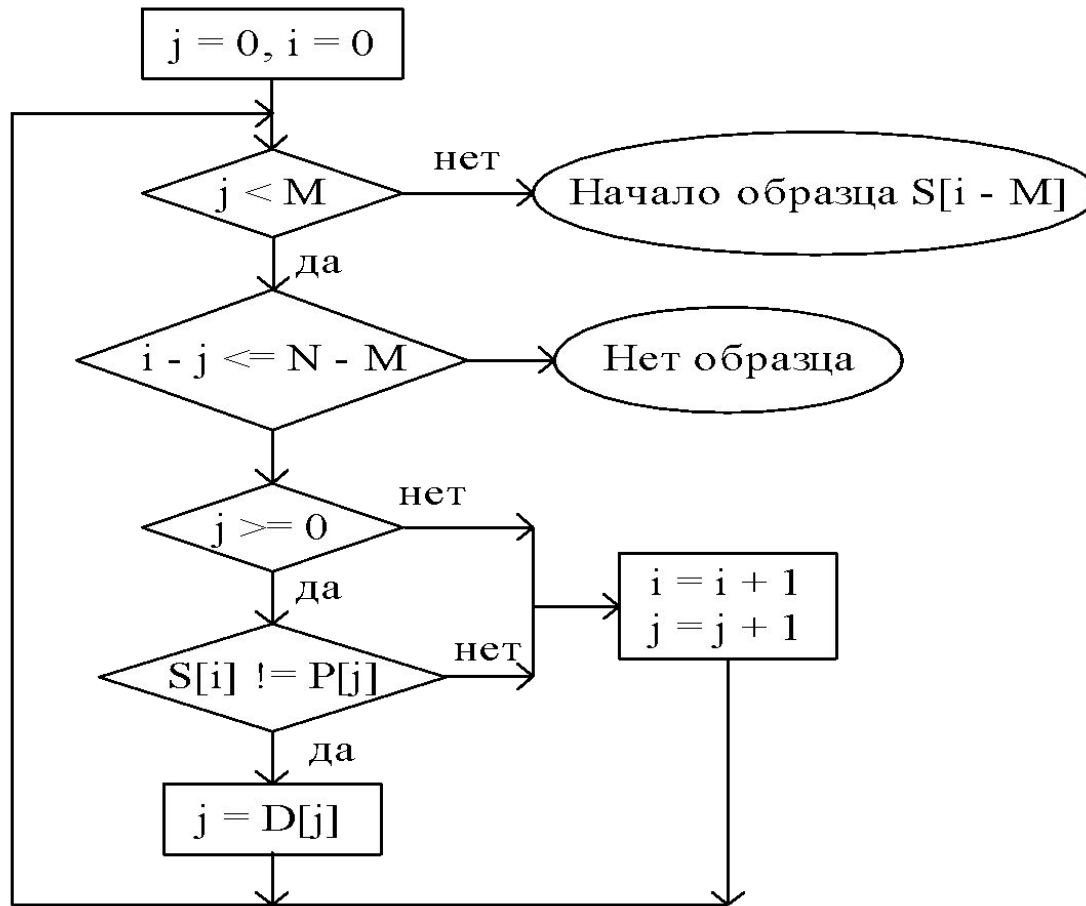
Образец и строка поиска могут быть двумя разными строками:



Более того, при таком подходе образец может не располагаться в начале строки:



Блок-схема КМП-поиска



i – индекс строки;

N – длина строки;

j – индекс подстроки;

M – длина подстроки.