

# Лекция 21. Обработка строк

Курс «Программирование»

КИТ, 2 семестр

Щукин Александр

Валентинович

# Алгоритмы поиска подстроки в строке

## 1. «Наивный» алгоритм

о б а    о б о б р а л и    о б о и    б о б р а  
о б о и

Число сравнений символов:

$$3 + 1 + 1 + 1 + 4 + 1 + 3 + 1 + 1 + 1 + 1 + 1 + 1 + 4 = 24$$

```
While ((Not Founded) And (i < H.Length - N.Length))  
    Founded = True  
    While ((j < N.Length) And (Founded))  
        Founded = (H(i + j) = N(j))  
        j = j + 1  
    End While  
    i = i + 1  
End While
```

Худший случай:      `simpleSearch("aaaaaaaaaaaaaaaaaaaaaaaaaab", "aaaaaab");`

## 2. Алгоритм Рабина - Карпа

о б а р л и

1 2 3 4 5 6 7

о	б	а		о	б	о	б	р	а	л	и		о	б	о	и		б	о	б	р	а
1	2	3	7	1	2	1	2	4	3	5	6	7	1	2	1	6	7	2	1	2	4	3

1	2	1	6
о	б	о	и

Функция:  $\sum_{i=1}^m s_i = 10$

Число сравнений символов:

0 + 0 + 0 + 0 + 0 + 0 + 3 + 0 + 0 + 0 + 0 + 0 + 0 + 4 = 7

Значения функции на подстроках:

13 13 13 11 6 9 10 14 18 21 19 16 11 10

# Алгоритм Кнута-Морриса-Пратта

- Префикс-функция (для строки)
- Префикс-функция строки  $\pi(S, i)$  – это длина наибольшего префикса строки  $S[1..i]$ , который не совпадает с этой строкой и одновременно является ее суффиксом. Проще говоря, это длина наиболее длинного начала строки, являющегося также и ее концом.

A	B	C	D	A	B	C	A	B	C	D	A	B	C	D	A	B
0	0	0	0	1	2	3	1	2	3	4	5	6	7	4	5	6
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6

### 3. Алгоритм Кнута - Морриса - Пратта

о	б	о	и
0	0	1	0

о	б	а		о	б	о	б	р	а	л	и		о	б	о	и		б	о	б	р	а
о	б	о	и																			

$$N' = N + (q - p(q))$$

а	б	р	а	к	а	д	а	б	р	а
0	0	0	1	0	1	0	1	2	3	4

а	б	р	а	к	а	д	а	б	р	а
0	0	0	1	0	1	0	1	2	3	4