

13.

Z-функция, построение за линейное

Время строки - количество
символов в этой строке,

время. Поиск подстроки в строке за
лин. время с использованием
Z-функции.

Сурфикс строки $A[j \dots |A|]$ - строка из $|A| - j + 1$ последних символов

Префикс строки $A[1 \dots i]$ - это строка из i первых символов строки A .

Сепаратор - $\$$

Z-функция - корреляция двух строк. Для строки S - массив Z_1, \dots, Z_n ,
такой, что Z_i - равен длине наибольшего общего префикса начинающегося
с позиции i сурфикса строки S и самой строки S .

(вектор длин наибольшего общего префикса строки с её сур-
фиксом)



Свойства префикса и сурфикса:

1) Пустая строка может быть использована
как префикс и как сурфикс

2) Любая подстрока является префиксом некотор. сурфикса и сур-
префиксом некотор. префикса

Для наглядности: строка ababcaaba

Хуффыне	строка	→	Z
<u>ababcaaba</u>	<u>ababcaaba</u>	→	(8)
babcaaba	ababcaaba	→	0
<u>abcaaba</u>	<u>ababcaaba</u>	→	(2)
bcaaba	ababcaaba	→	0
caaba	ababcaaba	→	0
<u>aaba</u>	<u>ababcaaba</u>	→	(3)
ba	ababcaaba	→	0
(a)	(a)babcaaba	→	(1)

сравниваем
на первом
символе
поки
и закончил
сходство

Для использования Z-функции для поиска подстроки в строке следует конкатенировать строку с текстом, разделив их символом:

ab\$ ababcaaba
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
17 0 0 1 0 2 0 4 0 0 4 0 0 1 0 2 0

→ кол-во вхождений

Z-блок - самая правая подстрока, равная подстроке.

префиксу, которую мы успешно обнаружили. Имеет границы left и right.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
A A B Z A A B Z A A B Z A A B Z A

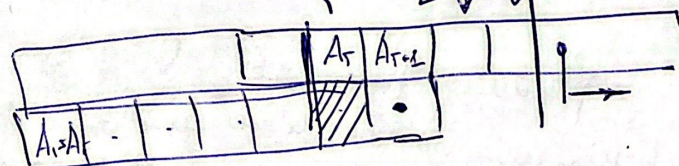
$$Z_p < |S(k...r)| \leq r - k + 1$$

$$1 < 7 - 5 + 8$$

$$1 < 3$$

(3) 0 0 1 1 0 0

r Z l



$$r = 7$$

$$i = 5$$

$$Z[i] = \min(7 - 5, Z[5 - 1])$$

$$Z[i] = \min(r - i, Z[i - 1])$$

наша новая граница.