

Pattern search: Z / KMP / Aho

Seminar 9.

What is a Z-function?

Definition

- For string S and position x
- Assume substring $P = S[x:]$
- **Z-func** - max. length of P 's prefix, being identical to the prefix of S
- Formally: $Z(S, x) = j$, where

$$j = \arg \max_i S[0 : i] = P[0 : i] = S[x : x + i]$$

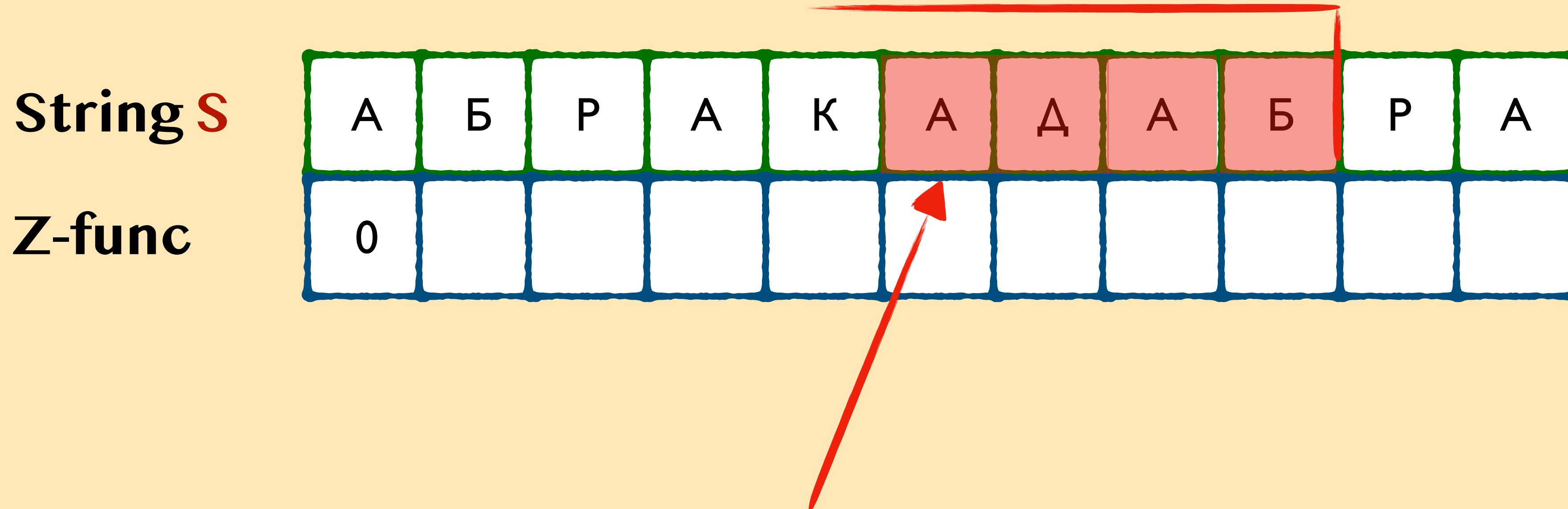
What is a Z-function?

String S

А Б Р А К А Д А Б Р А

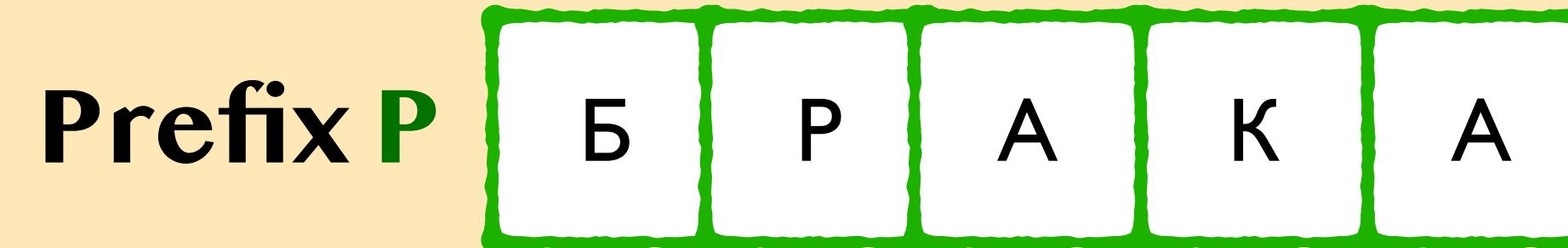
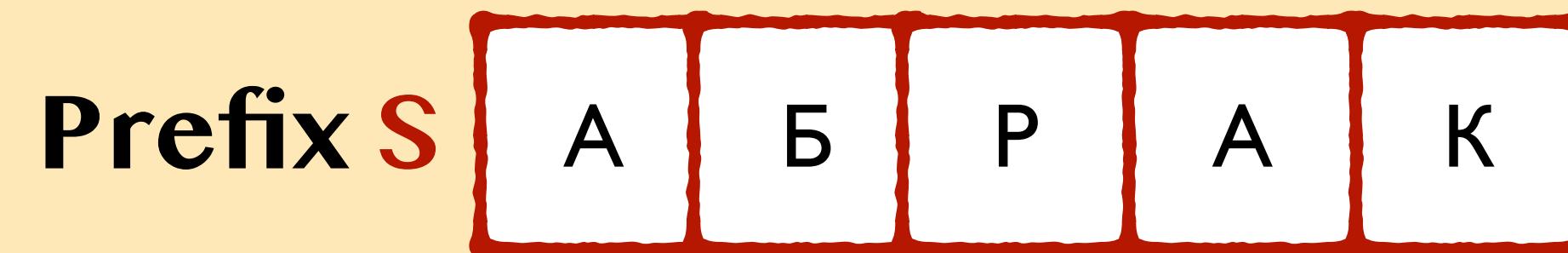
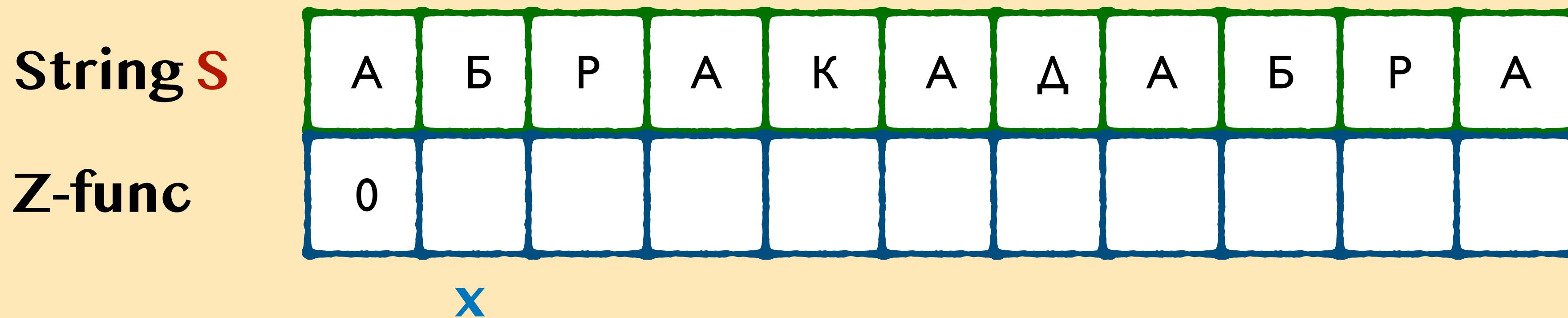
Z-func

What is a Z-function?

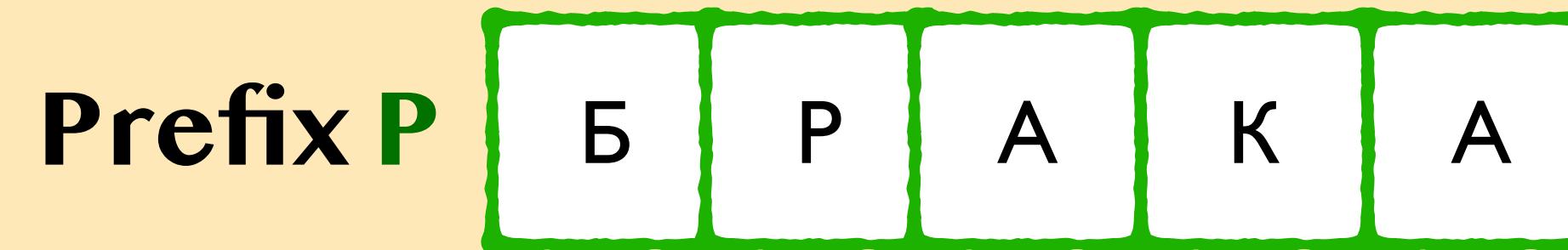
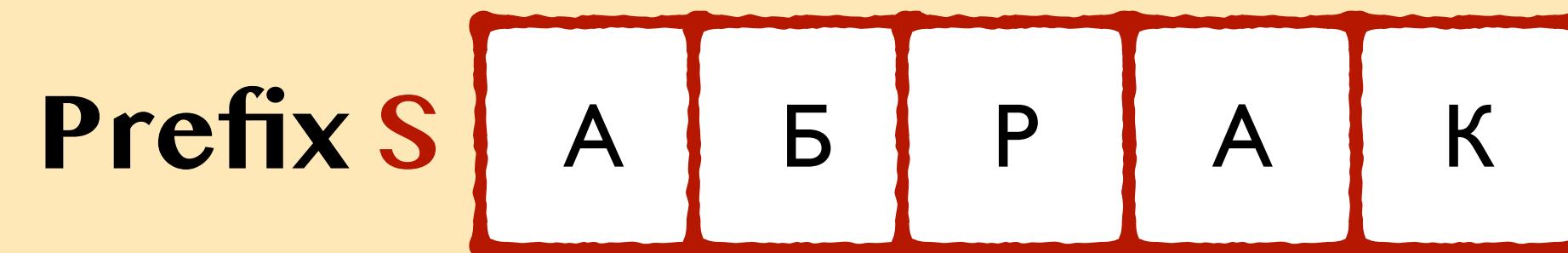
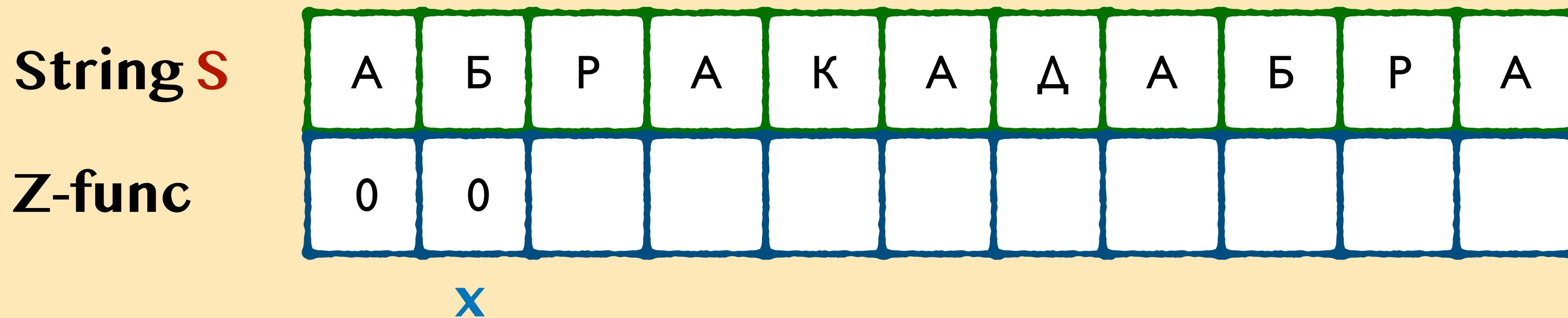


Prefix of size 4 for $S[5] = 'АДАБ'$

What is a Z-function?



What is a Z-function?



What is a Z-function?

String S

A	Б	Р	А	К	А	Д	А	Б	Р	А
0	0	0	I	0	I	0				

Z-func

X

Prefix S

A	Б	Р	А	К
---	---	---	---	---

Prefix P

A	Б	Р	А	
---	---	---	---	--

What is a Z-function?

String S	A	Б	Р	А	К	А	Д	А	Б	Р	А
Z-func	0	0	0	1	0	1	0	4	0	0	1

Prefix S	A	Б	Р	А	К
-----------------	---	---	---	---	---

Prefix P	A	Б	Р	А	
-----------------	---	---	---	---	--

What is a Z-function?

Pro Level

What is a Z-function?

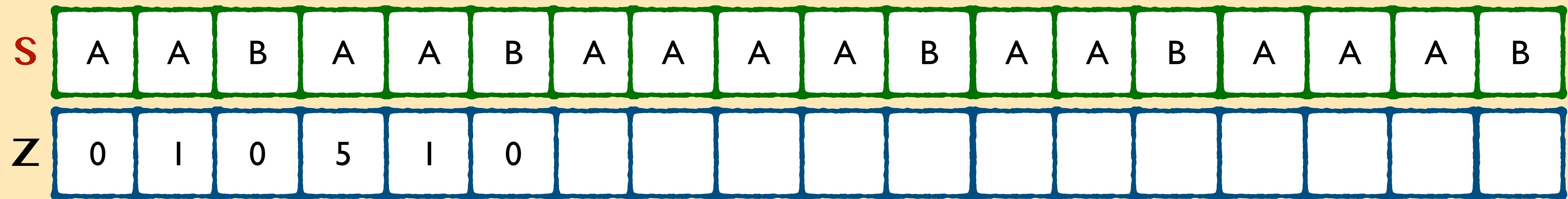
Pro Level

S	A	A	B	A	A	B	A	A	A	A	B	A	A	B	A	A	B	
Z	0	I	0	5	I	0	2	2	9	I	0	5	I	0	2	3	I	0

What is a Z-function?

Efficient Implementation

Idea — we want to avoid calculating already known values & start naive implementation from the best index

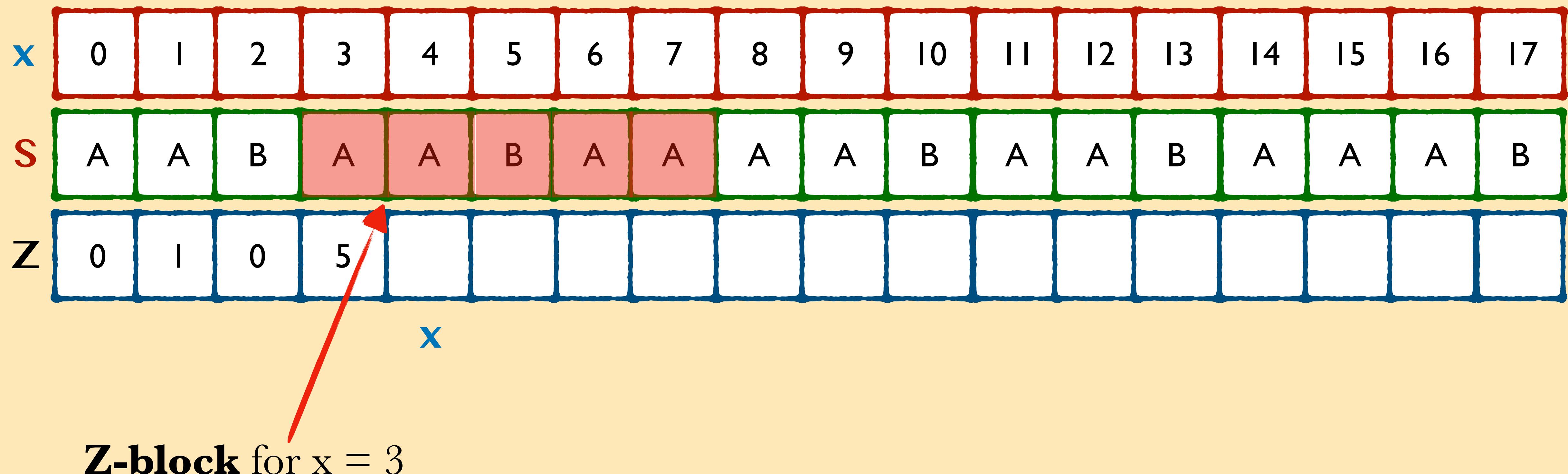


Definition

Z-block — substring $S[x : z[x] + x - 1]$

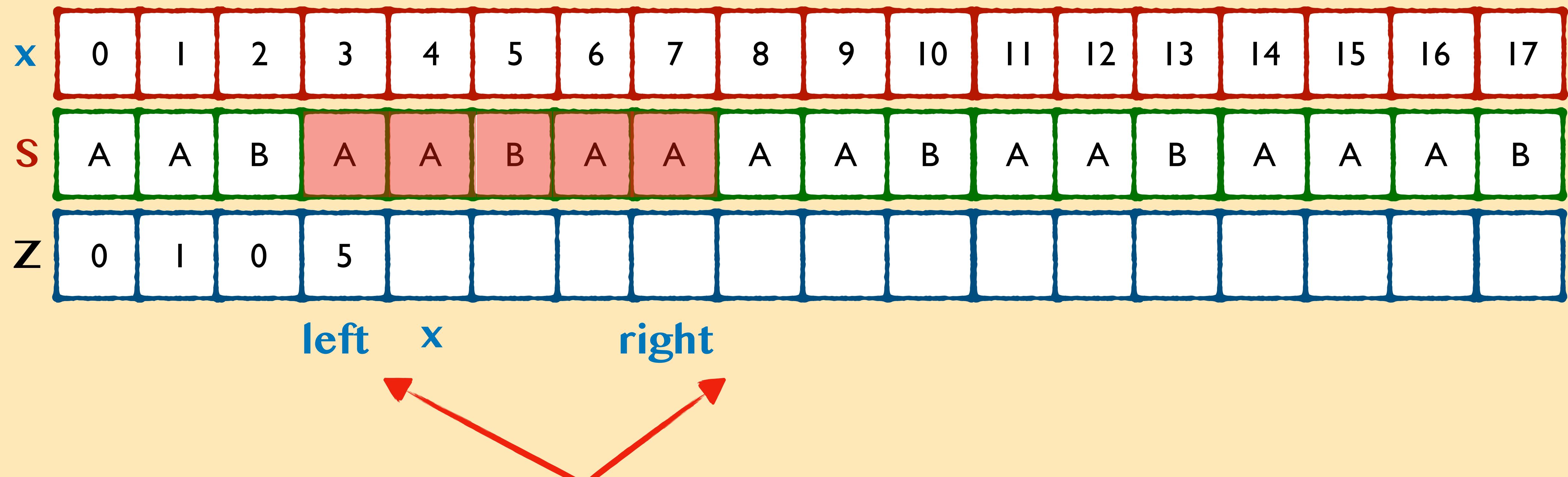
What is a Z-function?

Efficient Implementation



What is a Z-function?

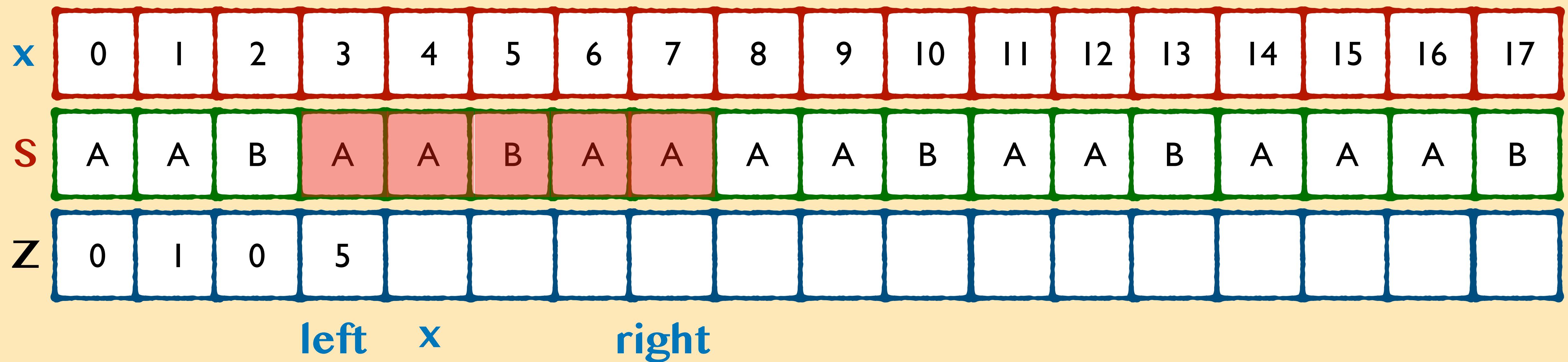
Efficient Implementation



left & right borders of Z-block with the greatest **right**

What is a Z-function?

Efficient Implementation



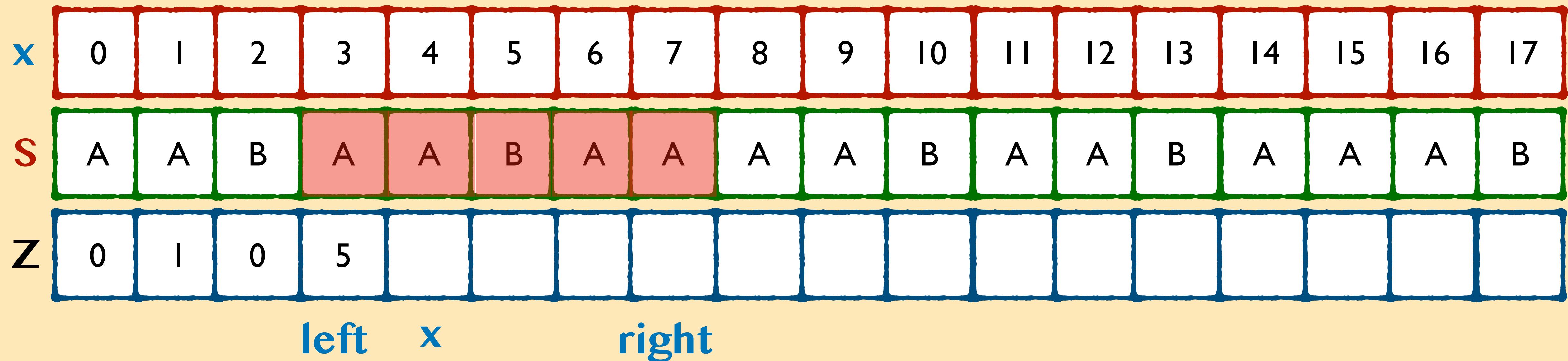
Case I: $x > \text{right}$ →

naive approach: $S[0:j] = P[0:j] = S[x:x+j]$

$$\begin{aligned} z[x] &= j - 1 \\ \text{left} &= x \\ \text{right} &= x + j - 1 \end{aligned}$$

What is a Z-function?

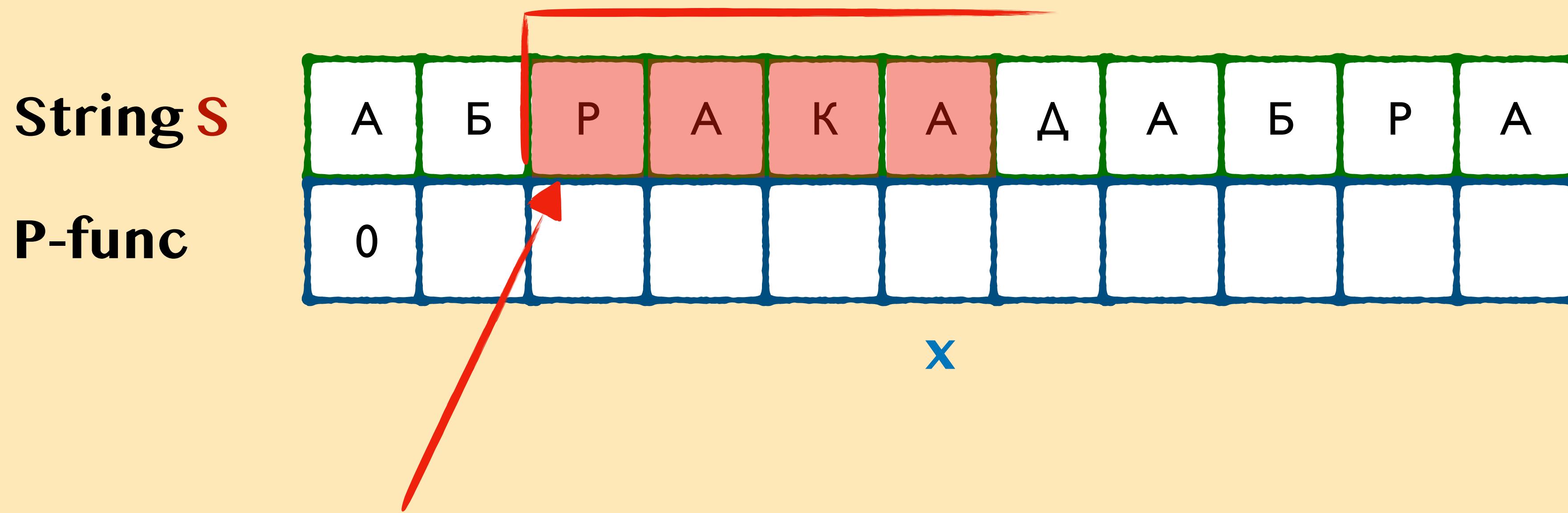
Efficient Implementation



Case II: $x \leq right \rightarrow z[x - left] ? right - x + 1$

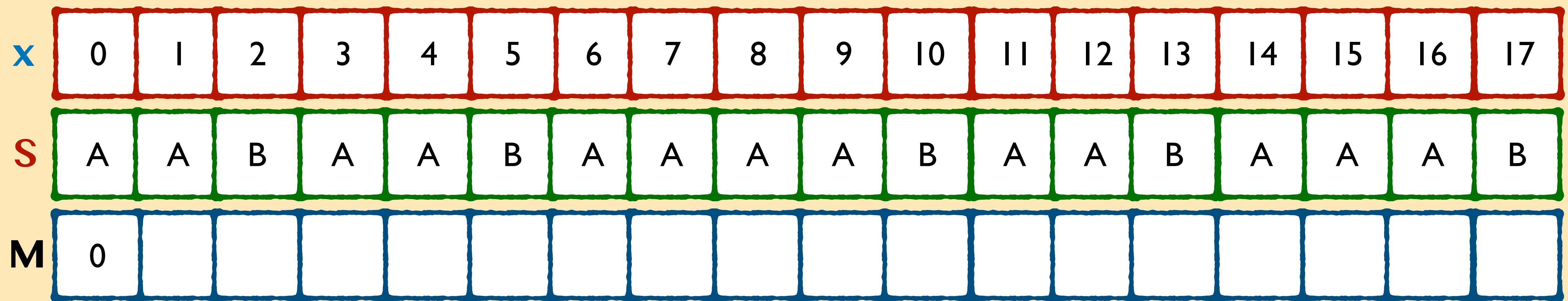
$\begin{cases} > & z_0[x] = right - x + 1 \rightarrow \text{naive approach} \\ \leq & z[x] = z[x-left] \end{cases}$

What is a prefix-function?



Suffix of size 4 for $S[5] = 'PAKA'$

What is a prefix-function?

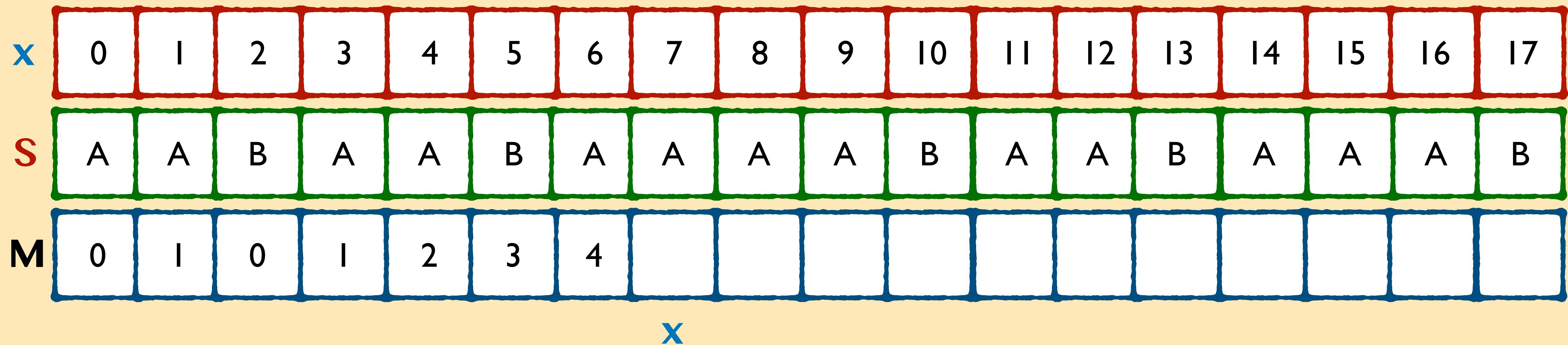


What is a prefix-function?

x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
s	A	A	B	A	A	B	A	A	A	A	B	A	A	B	A	A	B	
m	0	1	0	1	2	3	4	5	2	2	3	4	5	6	7	8	9	3

What is a prefix-function?

Efficient Implementation



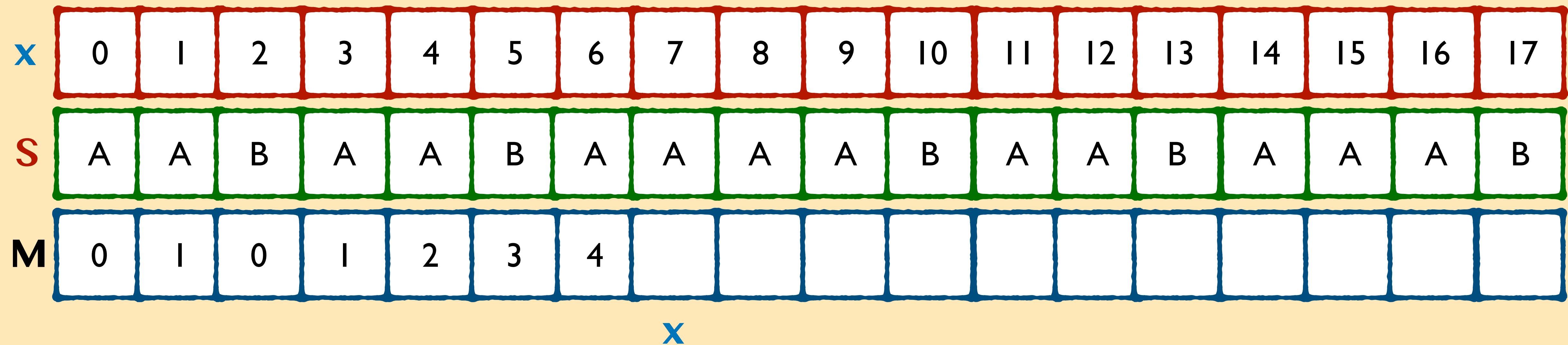
Обозначения

$S(x)$ - префикс S длины x

$P(x)$ - префикс и суффикс
максимальной длины для
позиции x

What is a prefix-function?

Efficient Implementation



Обозначения

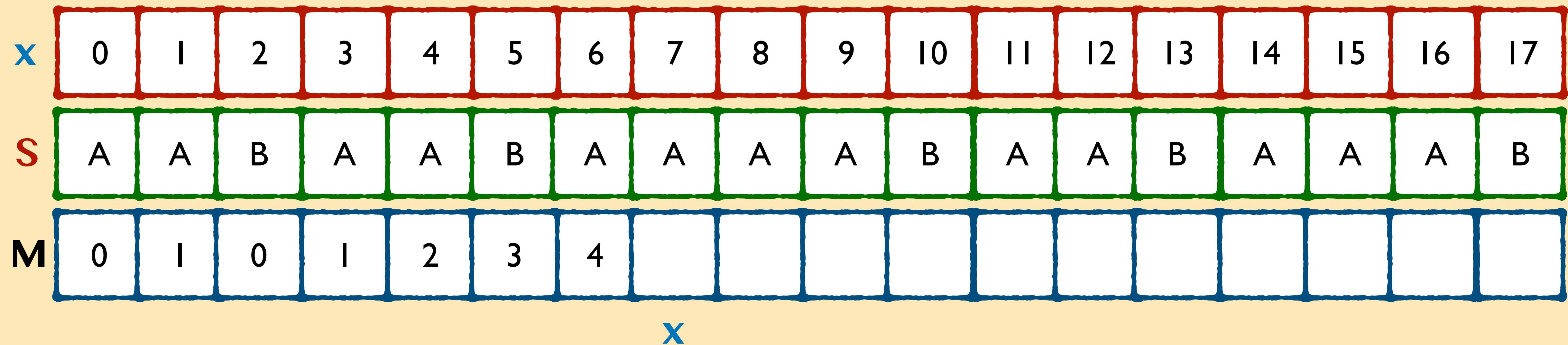
$S(x)$ - префикс S длины x

$P(x)$ - префикс и суффикс
максимальной длины для
позиции x

$$\longrightarrow S(M[x]) = P(x)$$

What is a prefix-function?

Efficient Implementation



Обозначения

$S(x)$ - префикс S длины x

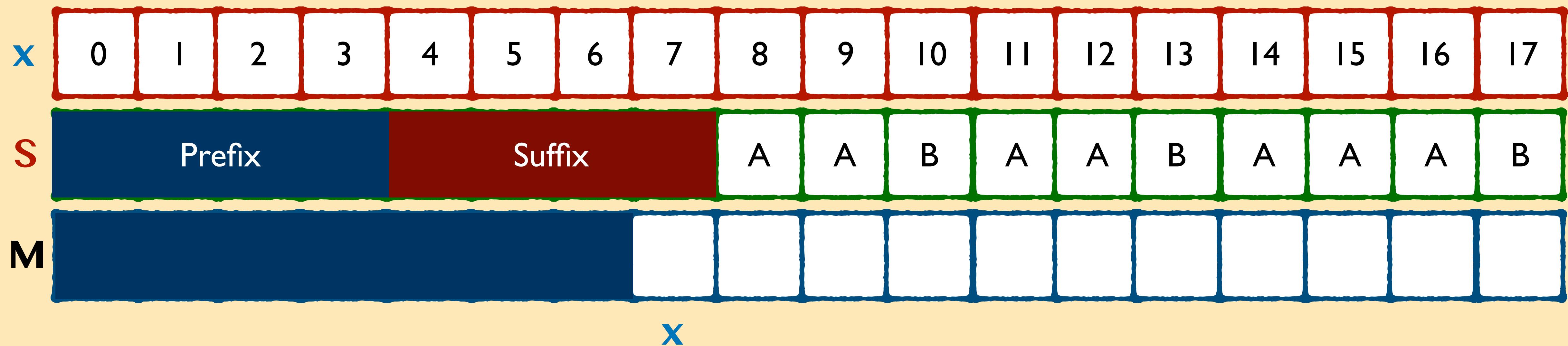
$P(x)$ - префикс и суффикс
максимальной длины для
позиции x



$$S(M[6]) = S(4) = 'AABA'$$

What is a prefix-function?

Efficient Implementation



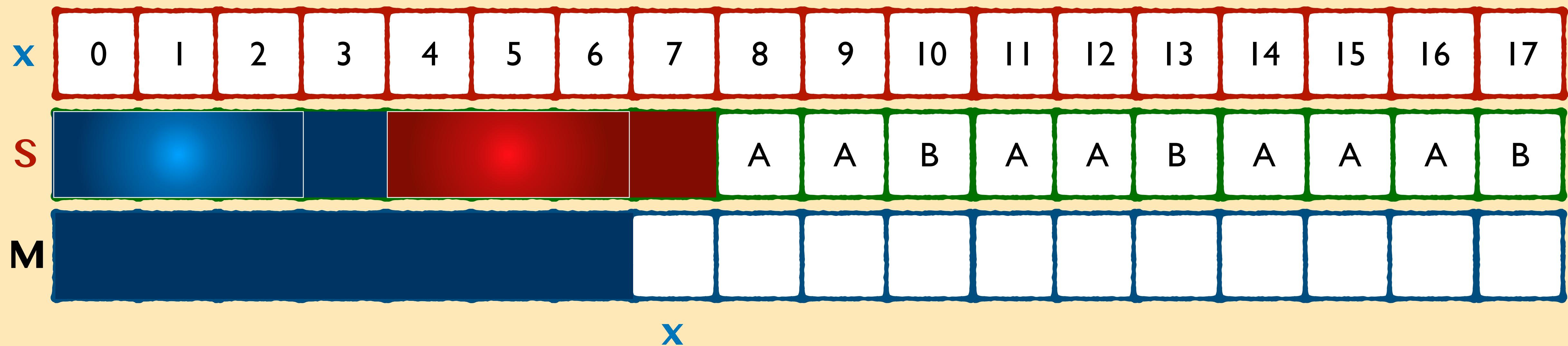
Мысль

Пусть $\exists P(7)$ длины k →

Например: $k = 4$

What is a prefix-function?

Efficient Implementation



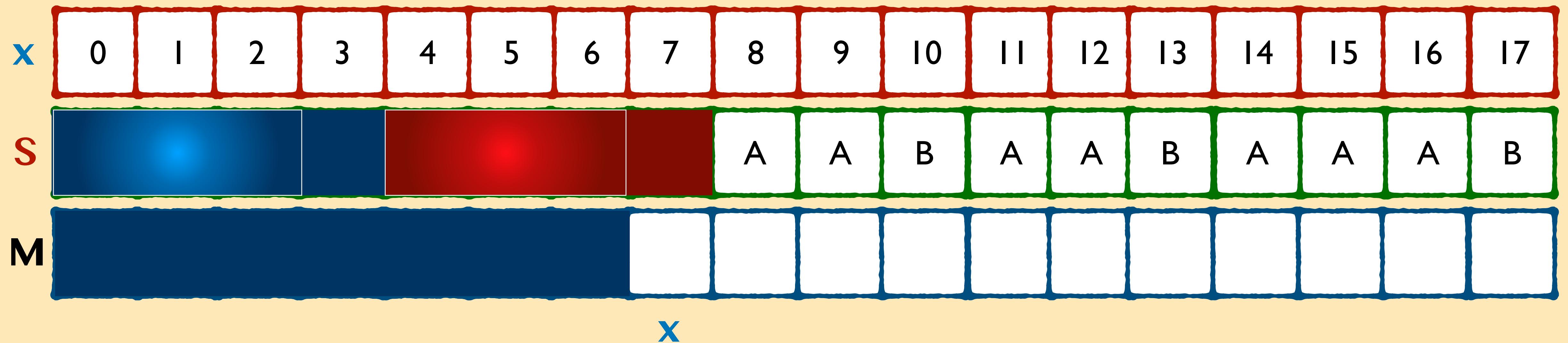
Мысль

Пусть $\exists P(7)$ длины k
Например: $k = 4$

→ Возьмем первые $k-1$ символов у $P(7) - p_6$

What is a prefix-function?

Efficient Implementation



Мысль
Кейс 1

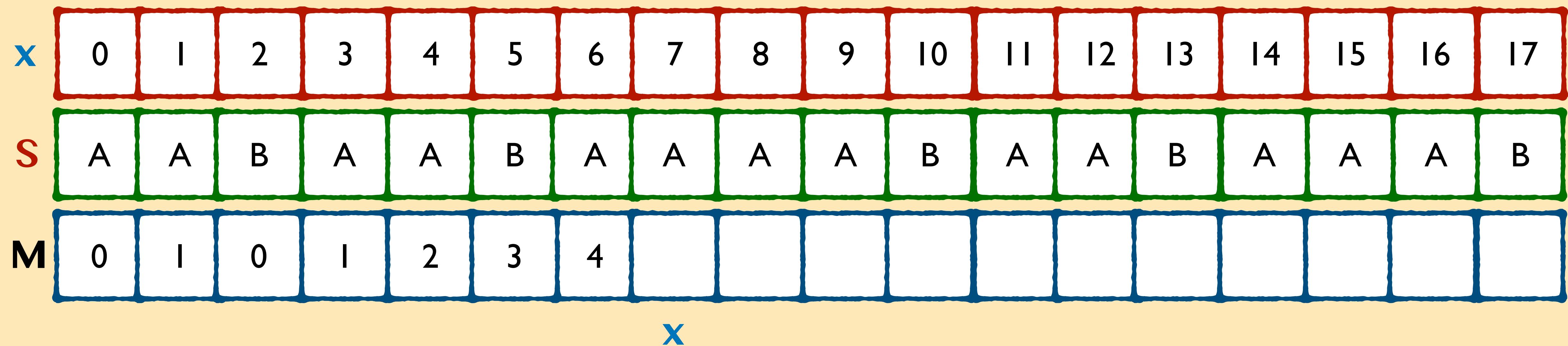
Пусть $\exists P(7)$ длины k
Например: $k = 4$

→ Возьмем первые $k-1$ символов у $P(7) - p_6$
если
 $p_6 = P(6)$ →
 $S[7] = S[M[6]]$

$$\begin{aligned}M[7] &= \\M[6] + 1 &\end{aligned}$$

What is a **prefix**-function?

Efficient Implementation

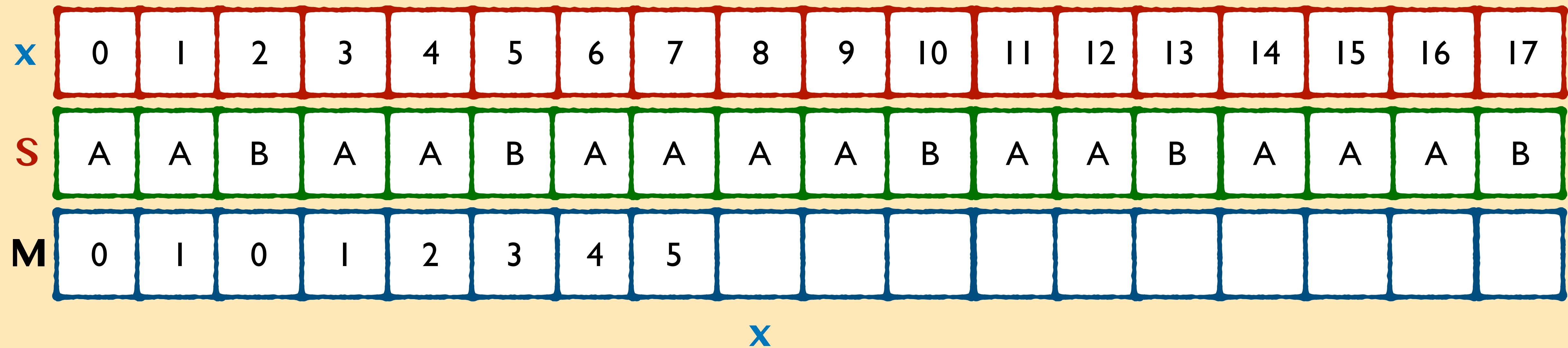


Мысль
Кейс 1

$$S[7] = 'A' = S[M[6]]$$

What is a prefix-function?

Efficient Implementation



Мысль
Кейс 2

если

$$\rightarrow p7 = P(7) \rightarrow$$

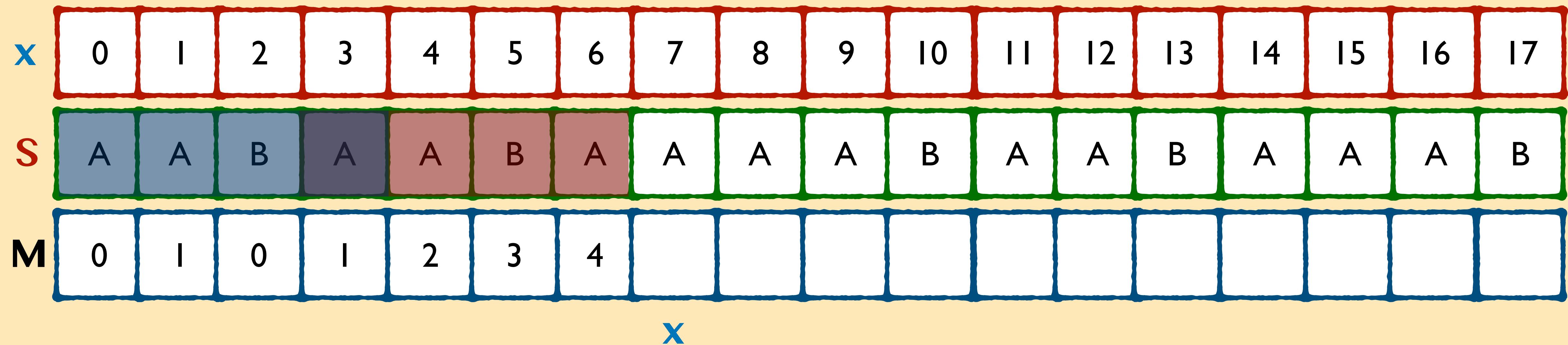
$$S[8] \neq S[M[7]]$$

Ищем более короткий
префикс с совпадающим
суффиксом

$S[M[M[x-1]-1]]$
до совпадения крайних
символов или 0 длины

What is a **prefix**-function?

Efficient Implementation



Обозначения

$$P(6) = \text{‘AABA’}$$

$P(x)$ - префикс и суффикс максимальной длины для позиции x