#### Algorytmy i struktury danych Wykład 4 wyszukiwanie wzorca w te<u>kście</u>

prof. dr hab. inż. Andrzej Obuchowicz

Instytut Sterowania i Systemów Informatycznych Uniwersytet Zielonogórski a.obuchowicz@issi.uz.zgora.pl p. 424 A2

26 października 2016

# Spis treści

- Wprowadzenie
- 2 Sformułowanie problemu
- Metoda naiwna
- Metoda Rabina-Karpa
- Metoda Knutha-Morrisa-Pratta
- Metoda Boyera-Moore'a
- Już za tydzień na wykładzie

# Sformułowanie problemu

problem wyszukiwania wzorca w tekscie

Dane: dany jest alfabet  $\Sigma$  zawierający  $|\Sigma| = b$  symboli, dany jest tekst (tablica) T długości n i wzorzec (tablica) W długości m;

*Szukane*: wszystkie podciągi tekstu *T* zgodne ze wzorcem *W*.

# Metoda naiwna (brute-force)

a	Ь	c	Ь	а	c	a	а	Ь	c	a	Ь	a	Ь	c	a	Ь	c	Ь	a
a	Ь	c	a	Ь															
a	Ь	c	Ь	a	c	a	a	Ь	c	a	Ь	a	Ь	С	a	Ь	c	Ь	a
	а	Ь	С	а	Ь														
а	ь	c	ь	а	c	а	а	ь	c	а	ь	а	ь	c	а	ь	c	ь	а
		а	Ь	С	а	ь													
а	ь	c	Ь	а	c	а	a	ь	С	а	ь	а	ь	c	а	Ь	c	Ь	а
			а	ь	С	а	ь												
а	ь	c	Ь	a	c	а	а		c	а	ь	а	ь	c	а	Ь	c	Ь	а
				a	Ь	c	а	ь											
а	ь	c	ь	а	c	а	а	ь		а	ь	а	ь	c	а	ь	c	ь	а
					a	ь	c	а	ь										
а	ь	c	Ь	а	c	a	a	ь	c	а	ь	а	ь	c	а	Ь	c	Ь	а
						a	Ь	c	а	ь									
а	Ь	c	ь	а	c	а	a	b	•	a	ь	а	ь	c	а	Ь	c	Ь	а
							a	ь	c	a	ь								
a	Ь	c	Ь	a	c	a	а	Ь	c	a	Ь	а	Ь	c	a	Ь	c	Ь	a
								a	Ь	c	а	ь							

liczba porównań elementów  $l_{por} = m(n-m+1) \sim O(nm)$ 

### Metoda Rabina-Karpa

$$k_{W} = W(1)b^{m-1} + W(2)b^{m-2} + \dots + W(m-1)b + W(m)$$
  

$$k_{T}(i) = T(i)b^{m-1} + T(i+1)b^{m-2} + \dots + T(i+m-2)b + T(i+m-1)$$
  

$$k_{T}(i+1) = (k_{T}(i) - T(i)b^{m-1})b + T(i+m)$$

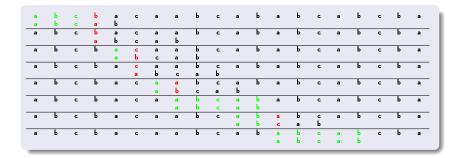
Wyznaczanie na podstawie powyższych wzorów - poblemy z reprezentacją wyniku dla alfabetów o dużej liczbie symboli i długich wzorców. Dlatego stosuje się arytmetykę modulo p, gdzie p jest pewną liczbą pierwszą.

liczba porównań elementów  $l_{por} = m(n-m+1) \sim O(nm)$ 

# Metoda Knutha-Morrisa-Pratta (KMP) - analiza wzorca

								przesunięcie o	kolejny element wzorca
a	b	С	a	b				1	0
	а	b	С	a	b				
a	b	С	a	b				1	0
	a	b	С	a	b				
a	Ь	С	а	b				2	0
		а	b	С	а	b			
a	Ь	С	a	b				3	0
			а	b	С	а	b		
a	Ь	С	a	Ь				3	1
			а	b	С	а	b		
a	Ь	С	a	Ь				3	2
			a	b	С	a	b		

#### Metoda KMP - przeszukiwanie

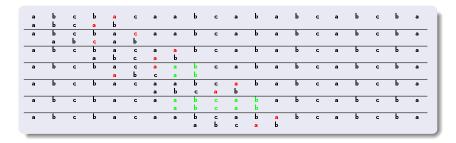


liczba porównań elementów  $I_{por} \sim O(m+n)$ 

# Metoda Boyera-Moore'a - analiza wzorca

wzorze	ec	a	b	С	a	b
		5	4	3	2	1
a		1	Х	2	1	!1
b		X	2	1	X	1
С		2	1	X	2	1
d		5	4	3	2	1
е		5	4	3	2	1

# Metoda Boyera-Moore'a - przeszukiwanie



liczba porównań elementów  $l_{por} \sim O(bm + n)$ 

## A w następnym tygodniu między innymi

#### Elementarne struktury danych:

- o to jest struktura danych,
- 2 co to jest słownik i zbiór liniowo uporządkowany,
- st os,
- kolejka,
- listy jedno i dwukierunkowe,
- listy cykliczne.

# Dziękuję za uwagę!!!