

# Algorytmy i struktury danych

## Wykład 4

### wyszukiwanie wzorca w tekście

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

- 1 Wprowadzenie
- 2 Sformułowanie problemu
- 3 Metoda naiwna
- 4 Metoda Rabina-Karpa
- 5 Metoda Knutha-Morrisa-Pratta
- 6 Metoda Boyera-Moore'a
- 7 Już za tydzień na wykładzie

# Sformułowanie problemu

problem wyszukiwania wzorca w tekście

*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	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a

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

# Metoda Rabina-Karpa

$$\begin{aligned}k_W &= W(1)b^{m-1} + W(2)b^{m-2} + \dots + W(m-1)b + W(m) \\k_T(i) &= T(i)b^{m-1} + T(i+1)b^{m-2} + \dots + T(i+m-2)b + T(i+m-1) \\k_T(i+1) &= (k_T(i) - T(i)b^{m-1})b + T(i+m)\end{aligned}$$

Wyznaczanie na podstawie powyższych wzorów - problemy 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	c	a	b				1	0
	a	b	c	a	b				
a	b	c	a	b				1	0
	a	b	c	a	b				
a	b	c	a	b				2	0
		a	b	c	a	b			
a	b	c	a	b				3	0
			a	b	c	a	b		
a	b	c	a	b				3	1
			a	b	c	a	b		
a	b	c	a	b				3	2
			a	b	c	a	b		

# Metoda KMP - przeszukiwanie

a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a

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

# Metoda Boyera-Moore'a - analiza wzorca

wzorzec	a	b	c	a	b
	5	4	3	2	1
a	1	x	2	1	!1
b	x	2	1	x	1
c	2	1	x	2	1
d	5	4	3	2	1
e	5	4	3	2	1



# Metoda Boyera-Moore'a - przeszukiwanie

a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	a	b	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a
a	b	c	b	a	c	a	a	b	c	a	b	a	b	c	a	b	c	b	a

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

# A w następnym tygodniu między innymi

## Elementarne struktury danych:

- 1 co to jest struktura danych,
- 2 co to jest słownik i zbiór liniowo uporządkowany,
- 3 stos,
- 4 kolejka,
- 5 listy jedno i dwukierunkowe,
- 6 listy cykliczne.

Dziękuję za uwagę!!!