

Algorytmy tekstowe

zajęcia 7.

Bartosz Górski, Tomasz Kulczyński, Błażej Osiński

Słowa

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

- słowo dla informatyka to ciąg znaków
- nie musi mieć sensu w żadnym używanym języku

Problemy

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Najczęściej stawiany problem tekstowy:

wyszukiwanie wzorca w tekście

- znak
- alfabet – zbiór dostępnych znaków
- słowa – ciągi znaków

Problemy

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Najczęściej stawiany problem tekstowy:

wyszukiwanie wzorca w tekście

- znak
- alfabet – zbiór dostępnych znaków
- słowa – ciągi znaków

Problemy

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Najczęściej stawiany problem tekstowy:

wyszukiwanie wzorca w tekście

- znak
- alfabet – zbiór dostępnych znaków
- słowa – ciągi znaków

Problemy

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Najczęściej stawiany problem tekstowy:

wyszukiwanie wzorca w tekście

- znak
- alfabet – zbiór dostępnych znaków
- słowa – ciągi znaków

Przykład

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Problem

- alfabet: $\{a, b\}$
- tekst: *abbbaabaabba*
- wzorzec: *aaba*

Przykład

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Problem

- alfabet: $\{a, b\}$
- tekst: *abbbaabaabba*
- wzorzec: *aba*

Rozwiązanie

- wystąpienia:
 - 1 *abbbaaba***aba***abaa*
 - 2 *abbbaab***aba***aa*

Pomysł

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

Spróbujmy w każdym miejscu dopasować wzorzec.

Działanie

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

*cbadab**cb**aabccdba*

bccd

***b**ccd*

bccd

bccd

bccd

***b**ccd*

Złożoność

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

n – długość tekstu
 m – długość wzorca

$$O(n \cdot m)$$

Złożoność

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

n – długość tekstu
 m – długość wzorca

$$O(n \cdot m)$$

W praktyce **znacznie** lepiej!

Czy rzeczywiście $O(n \cdot m)$?

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

- tekst: $aaaaaa \dots aaaa$ (długości n)
- wzorzec: $aa \dots aab$ (długości m)
- ile operacji wykonuje algorytm naiwny?

Czy rzeczywiście $O(n \cdot m)$?

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

- tekst: $aaaaaa \dots aaaa$ (długości n)
- wzorzec: $aa \dots aab$ (długości m)
- liczba operacji: $(n - m + 1) \cdot m$

Pomysł

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

Za każdym razem zaczynamy porównywać od początku –
tracimy cenne informacje!

Pomysł

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica *P*

Za każdym razem zaczynamy porównywać od początku –
tracimy cenne informacje!

Przykład

*ab***aaa**aaaaaaaaa
aaa*b*

Pomysł

Algorytmy tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Za każdym razem zaczynamy porównywać od początku –
tracimy cenne informacje!

Przykład

*ab***aaa**aaaaaaaaa
aaa*b*

*aba***aa**aaaaaaaaa
aa*ab*

Trudne słowa

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica *P*

- prefiks
- sufiks
- prefikso-sufiks

Trudne słowa

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

- prefiks
- sufiks
- prefikso-sufiks

Trudne słowa

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

- prefiks
- sufiks
- prefikso-sufiks

Tablica najdłuższych prefikso-sufiksów – $P[]$

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

$P[i]$ – długość najdłuższego właściwego prefikso-sufiksu
prefiksu długości i słowa w

Tablica najdłuższych prefikso-sufiksów – $P[]$

Algoritmy
tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm
naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

$P[i]$ – długość najdłuższego właściwego prefikso-sufiksu
prefiksu długości i słowa w

słowo w		b	a	a	b	a	b	a	a
$i =$	0	1	2	3	4	5	6	7	8
$P[i] =$	0	0	0	0	1	2	1	2	3

Obliczanie $P[]$

Algoritmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

$P[]$	0	0	1	1	2	3			
słowo	a	b	a	a	b	a	b	a	a
pasujący prefikso-sufiks				a	b	a			

Obliczanie $P[]$

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

$P[]$
słowo
pasujący prefikso-sufiks
kolejna pozycja

0	0	1	1	2	3				
a	b	a	a	b	a	b	a	a	
			a	b	a				
			a	b	a	a			

Obliczanie $P[]$

Algorytmy tekstowe

Wstęp

Wprowadzenie
Przykład

Algorytm naiwny

Rozwiązanie
Czas

Algorytm KMP

Tablica P

$P[]$
słowo
pasujący prefikso-sufiks
kolejna pozycja

0	0	1	1	2	3	2		
a	b	a	a	b	a	b	a	a
			a	b	a			
			a	b	a	a		
					a	b		

Wyszukiwanie wzorca

Algoritmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P [:

P [:		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
-------	--	---	---	---	---	---	---	---	---	---	---	---

Wyszukiwanie wzorca

Algoritmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P []:

P []		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
		a	b									

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P [:

P [:		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
				a	b							

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P []:

P []		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
				a	b	a						

Wyszukiwanie wzorca

Algoritmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P []:

P []		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
			a	b	a							

Wyszukiwanie wzorca

Algoritmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P [:

P [:	0	0	1	1	2	3
słowo	a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst	a	a	b	a	b	a	a	b	a	a	a
		a	b	a	a						

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P []:

P []		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
					a	b						

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P []:

P []		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
					a	b	a					

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P []:

P []		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
					a	b	a	a				

Wyszukiwanie wzorca

Algoritmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P []:

P []		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
					a	b	a	a	b			

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica $P[]$:

$P[]$		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
					a	b	a	a	b	a		

Trafienie! *aab**aba**baaa*

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P [:]

P [:]	0	0	1	1	2	3
słowo	a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst	a	a	b	a	b	a	a	b	a	a	a
							a	b	a		

Trafienie! *aab**aba**baaa*

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica $P[[]]$:

$P[[]]$		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
							a	b	a	a		

Trafienie! *aab***aba***abaaa*

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica $P[]$:

$P[]$		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
							a	b	a	a	b	

Trafienie! *aab**aba**abaaa*

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica $P[]$:

$P[]$		0	0	1	1	2	3
słowo		a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst		a	a	b	a	b	a	a	b	a	a	a
											a	b

Trafienie! *aab**aba**baaa*

Wyszukiwanie wzorca

Algorytmy
tekstowe

Wstęp

Wprowadzenie

Przykład

Algorytm
naiwny

Rozwiązanie

Czas

Algorytm KMP

Tablica P

Wzorzec i jego tablica P [:

P [:	0	0	1	1	2	3
słowo	a	b	a	a	b	a

Wyszukiwanie wzorca:

tekst	a	a	b	a	b	a	a	b	a	a	a
											a

Trafienie! *aab**aba**baaa*