

# Algoritmi i strukture podataka

10 Obrada teksta

Katedra za informatiku, Fakultet tehničkih nauka, Novi Sad

2024

# Osnovni pojmovi

- neka je  $P$  string dužine  $m$ 
  - podstring  $P [i..j]$  od  $P$  je podsekvenca od  $P$  koja sadrži karaktere sa rangom između  $i$  i  $j$
  - prefiks od  $P$  je podstring tipa  $P [0..i]$
  - sufiks od  $P$  je podstring tipa  $P [i..m - 1]$

# Boyer-Moore – primer algoritma

Šablon: a b c a b

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

Tekst: a b a a b c c b d a b c a b b b a b  
          \x | |  
          a b c a b

Karakter šablonu	a	b	c
Vrednost funkcije	3	4	2

- Uzmimo da je  $m$  dužina šablonu.
- Poređenje kreće unazad. Razlika je na karakteru  $a$  teksta na indeksu  $i=2$ , indeks šablonu  $k=2$ . Iščitavamo vrednost iz tabele za karakter  $a$ . Nailazimo na vrednost 3 što je veće od trenutnog  $k$ .
- Ako bismo pomerali šablon tako da se  $a$  iz teksta poklopi sa poslednjim pojavljivanjem karaktera  $a$  u tekstu, šablon bismo pomerali uлево što je neprihvatljivo.

# Boyer-Moore – primer algoritma

Šablon: a b c a b

Karakter šablonu	a	b	c
Vrednost funkcije	3	4	2

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

Tekst: a b a a b c c b d a b c a b b b a b  
           $\ddagger$  | |  
          a b c a b

- U ovom slučaju, pomeramo šablon za jedno mesto udesno (kao u brute force algoritmu).
- Resetujemo indeks šablonu na kraj,  $k=4$  (uvek  $m-1$ ).
- Indeks teksta se vraća prvo za broj pozicija koliko se pomerio od kraja šablonu pa zatim za još jednu poziciju ( $m-k$ )

# Boyer-Moore – primer algoritma

Šablon: a b c a b

Karakter šablonu	a	b	c
Vrednost funkcije	3	4	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b a c b d a b c a b b b a b  
                  |  
          a b c a b

- Razlika na karakteru *a* teksta na indeksu  $i=5$ , indeks šablonu  $k=4$ . Iščitavamo vrednost iz tabele za karakter *a*. Nailazimo na vrednost 3 što je manje od trenutnog  $k$ . Treba da poravnamo *a* iz teksta sa poslednjim pojavljivanjem *a* iz šablonu.
- Indeks u tekstu treba da povećamo za 1 (da dobijemo poravnanje).
- Treba da odredimo formulu po kojoj određujemo povećanje  $i$ :

$i += m - vrednost\_iz\_tabele - 1$

- Promenljivu  $k$  resetujemo na kraj,  $k = 4$

# Boyer-Moore – primer algoritma

Šablon: a b c a b

Karakter šablonu	a	b	c
Vrednost funkcije	3	4	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b a c b d a b c a b b b a b  
                  |  
          a b c a b

- Razlika na karakteru  $c$  u tekstu na indeksu  $i=6$ , indeks šablonu  $k=4$ . Iščitavamo vrednost iz tabele za karakter  $c$ . Nailazimo na vrednost 2 što je manje od trenutnog  $k$ . Treba da poravnamo  $c$  iz teksta sa poslednjim pojavljivanjem  $c$  iz šablonu.
- Indeks u tekstu treba da povećamo za 2 (da dobijemo poravnanje).
- Korigujemo i po formuli od malopre.

$$i += m - \text{vrednost\_iz\_tabele} - 1 = 5 - 2 - 1 = 2$$

- Promenljivu  $k$  resetujemo na kraj,  $k = 4$

# Boyer-Moore – primer algoritma

Šablon: a b c a b

Karakter šablonu	a	b	c
Vrednost funkcije	3	4	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b a c b d a b c a b b b a b  
                  |  
          a b c a b

- Razlika je na karakteru  $d$  teksta na indeksu  $i=8$ , indeks šablonu  $k=4$ . Iščitavamo vrednost iz tabele za karakter  $d$ . Pošto se karakter  $d$  ne nalazi u šablonu, uzimamo da je vrednost funkcije -1. Šablon treba da pomerimo za čitavu dužinu.
- Korigujemo i po formuli od malopre.

$$i += m - \text{vrednost\_iz\_tabele} - 1 = 5 + 1 - 1 = 5$$

- Promenljivu  $k$  resetujemo na kraj,  $k = 4$

# Boyer-Moore – primer algoritma

Šablon: a b c a b

Karakter šablonu	a	b	c
Vrednost funkcije	3	4	2

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

Tekst: a b a a b a c b d a b c a b b b a b

| | | | |  
a b c a b

- Svi karakteri šablonu su se poklopili sa karakterima teksta.
- Postupak je završen.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8  
Šablon: b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablona	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	0	0	0	0	0	0

- Na početku, sve vrednosti tabele se inicijalizuju na 0.
- Za računanje vrednosti funkcije za poziciju n, gledamo dužinu prefiksa i sufiksa uključujući i poziciju n (dakle, od 0 do n)
- Preskačemo računanje za poziciju 0 (svakako je 0).
- Krećemo od pozicije 1.

# KMP – primer funkcije neuspeha

Šablon:      0 1 2 3 4 5 6 7 8  
                b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	0	0	0	0	0	0

- Za određivanje maksimalne dužine istog prefiksa i sufiksa, uvodimo dve promenljive **k** i **i**. Promenljiva **i** odgovara poziciji trenutnog karaktera, a **k** broj poklapanja. Na početku, **i** dobija vrednost indeksa za koji računamo vrednost funkcije.
- Na početku:
$$\begin{aligned} k &= 0 \\ i &= 1 \end{aligned}$$
- Poredimo vrednosti elemenata na pozicijama **k** i **i**. U pitanju su karakteri *b* i *a*. Zaključujemo da su različiti, računanje za poziciju 1 se prekida, ostaje upisana 0. Prelazimo na sledeću poziciju.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8  
Šablon: b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	0	0	0	0	0	0

$$k = 0$$

$$i = 2$$

- Poredimo vrednosti elemenata na pozicijama **k** i **i**. U pitanju su karakteri *b* i *a*. Zaključujemo da su različiti, računanje za poziciju 2 se prekida, ostaje upisana 0. Prelazimo na sledeću poziciju.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8  
Šablon: b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	0	0	0	0	0	0

$$\begin{aligned} k &= 0 \\ i &= 3 \end{aligned}$$

- Poredimo vrednosti elemenata na pozicijama **k** i **i**. U pitanju su karakteri *b* i *b*. Zaključujemo da su jednaki (vrednost funkcije se povećava sa 0 na 1). Ta vrednost se upisuje za poziciju  $i=3$ .
- Nastavljamo. Proverićemo da li postoji duži prefiks jednak sufiksu.
- Uvećavamo **k** i **i** ( $k=1$ ,  $i=4$ ).

# KMP – primer funkcije neuspeha

Šablon:      0 1 2 3 4 5 6 7 8  
                b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	1	0	0	0	0	0

$k = 1$   
 $i = 4$



- Poredimo vrednosti elemenata na pozicijama  $k$  i  $i$ . U pitanju su karakteri  $a$  i  $b$  koji se ne poklapaju. Dakle, niz poklapanja se ovde prekinuo. Ostaje da proverimo da li postoji kraći niz prefiksa jednak sufiksima. Da bismo to proverili,  $k$  menja vrednost i dobija vrednost funkcije prethodnog odnosno  $(k-1)$ -og elementa.

$k = 0$   
 $i = 4$

- Poredimo vrednosti elemenata na pozicijama  $k$  i  $i$ . U pitanju su karakteri  $b$  i  $b$  koji se poklapaju. Promenljiva  $k$  se uvećava za 1. Nova vrednost  $k$  se upisuje kao vrednost funkcije na poziciji  $i$  (odnosno 4). I promenljiva  $i$  se uvećava za 1.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8  
Šablon: b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	1	1	0	0	0	0

$k = 1$   
 $i = 5$

- Poredimo vrednosti elemenata na pozicijama  $k$  i  $i$ . U pitanju su karakteri  $a$  i  $a$  koji se poklapaju. Dakle, niz poklapanja se ovde nastavlja. Promenljiva  $k$  se uvećava za 1. Nova vrednost  $k$  se upisuje kao vrednost funkcije na poziciji  $i$  (odnosno 5). I promenljiva  $i$  se uvećava za 1.

$k = 2$   
 $i = 6$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8  
Šablon: b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	1	1	2	0	0	0

$$k = 2$$

$$i = 6$$

- Poredimo vrednosti elemenata na pozicijama **k** i **i**. U pitanju su karakteri *a* i *a* koji se poklapaju. Dakle, niz poklapanja se ovde nastavlja. Promenljiva **k** se uvećava za 1. Nova vrednost **k** se upisuje kao vrednost funkcije na poziciji **i** (odnosno 6). I promenljiva **i** se uvećava za 1.

$$k = 3$$

$$i = 7$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8  
Šablon: b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	1	1	2	3	0	0

$$k = 3$$

$$i = 7$$

- Poredimo vrednosti elemenata na pozicijama **k** i **i**. U pitanju su karakteri *b* i *b* koji se poklapaju. Dakle, niz poklapanja se ovde nastavlja. Promenljiva **k** se uvećava za 1. Nova vrednost **k** se upisuje kao vrednost funkcije na poziciji **i** (odnosno 7). I promenljiva **i** se uvećava za 1.

$$k = 4$$

$$i = 8$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8  
Šablon: b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	1	1	2	3	4	0

$$\begin{aligned} k &= 4 \\ i &= 8 \end{aligned}$$



- Poredimo vrednosti elemenata na pozicijama  $k$  i  $i$ . U pitanju su karakteri  $b$  i  $a$  koji se poklapaju. Dakle, niz poklapanja se ovde prekida. Ostaje da proverimo da li postoji kraći niz prefiksa jednak sufiksima. Da bismo to proverili,  $k$  menja vrednost i dobija vrednost funkcije prethodnog odnosno  $(k-1)$ -og elementa.

$$\begin{aligned} k &= 1 \\ i &= 8 \end{aligned}$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8  
Šablon: b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	1	1	2	3	4	0

$$k = 1$$

$$i = 8$$

- Poredimo vrednosti elemenata na pozicijama **k** i **i**. U pitanju su karakteri *a* i *a* koji se poklapaju. Dakle, niz poklapanja se ovde nastavlja. Promenljiva **k** se uvećava za 1. Nova vrednost **k** se upisuje kao vrednost funkcije na poziciji **i** (odnosno 8). I promenljiva **i** se uvećava za 1.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8  
Šablon: b a a b b a a b a

Indeks	0	1	2	3	4	5	6	7	8
Karakter šablonu	b	a	a	b	b	a	a	b	a
Vrednost funkcije	0	0	0	1	1	2	3	4	2

- Promenljiva i se uvećala na vrednost 9 čime izlazi iz opsega dostupnih indeksa.
- Postupak se ovde prekida.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7  
Šablon: a b c d a b c d

Indeks	0	1	2	3	4	5	6	7
Karakter šablonu	a	b	c	d	a	b	c	d
Vrednost funkcije	0	0	0	0	0	0	0	0

$$k = 0$$

$$i = 1$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7  
Šablon: a b c d a b c d

Indeks	0	1	2	3	4	5	6	7
Karakter šablonu	a	b	c	d	a	b	c	d
Vrednost funkcije	0	0	0	0	0	0	0	0

$$k = 0$$

$$i = 2$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7  
Šablon: a b c d a b c d

Indeks	0	1	2	3	4	5	6	7
Karakter šablonu	a	b	c	d	a	b	c	d
Vrednost funkcije	0	0	0	0	0	0	0	0

$$k = 0$$

$$i = 3$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7  
Šablon: a b c d a b c d

Indeks	0	1	2	3	4	5	6	7
Karakter šablonu	a	b	c	d	a	b	c	d
Vrednost funkcije	0	0	0	0	0	0	0	0

$$k = 0$$

$$i = 4$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=4**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7  
Šablon: a b c d a b c d

Indeks	0	1	2	3	4	5	6	7
Karakter šablonu	a	b	c	d	a	b	c	d
Vrednost funkcije	0	0	0	0	1	0	0	0

$$k = 1$$

$$i = 5$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=5**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7  
Šablon: a b c d a b c d

Indeks	0	1	2	3	4	5	6	7
Karakter šablonu	a	b	c	d	a	b	c	d
Vrednost funkcije	0	0	0	0	1	2	0	0

$$k = 2$$

$$i = 6$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=6**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7  
Šablon: a b c d a b c d

Indeks	0	1	2	3	4	5	6	7
Karakter šablonu	a	b	c	d	a	b	c	d
Vrednost funkcije	0	0	0	0	1	2	3	4

$$k = 3$$

$$i = 7$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=7**. Promenljiva **i** se povećava.
- Kraj postupka.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	0	0	0	0	0	0	0	0	0	0	0

$$k = 0$$

$$i = 1$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=1**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	0	0	0	0	0	0	0

$$k = 1$$

$$i = 2$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=2**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	0	0	0	0	0	0	0

$$k = 2$$

$$i = 3$$

- Nema poklapanja. Pošto je  $k > 0$  znamo da je pre toga bilo poklapanja. Promenljiva  $k$  dobija vrednost funkcije na poziciji  $k-1$ , dakle,  $k$  dobija vrednost 1.

$$k = 1$$

$$i = 3$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	0	0	0	0	0	0	0



$$k = 1$$

$$i = 3$$

- Nema poklapanja. Pošto je  $k > 0$  znamo da je pre toga bilo poklapanja. Promenljiva  $k$  dobija vrednost funkcije na poziciji  $k-1$ , dakle,  $k$  dobija vrednost 0.

$$k = 0$$

$$i = 3$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablona	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	0	0	0	0	0	0	0

$$k = 0$$

$$i = 3$$

- Nema poklapanja. Pošto je  $k=0$  zaključujemo da ne postoji prefiks jednak sufiksu.
- Uvećavamo  $i$ .

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablona	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	0	0	0	0	0	0	0

$$k = 0$$

$$i = 4$$

- Nema poklapanja. Pošto je  $k=0$  zaključujemo da ne postoji prefiks jednak sufiksu.
- Povećavamo  $i$ .

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	0	0	0	0	0	0	0

$$k = 0$$

$$i = 5$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=5**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	1	0	0	0	0	0	0

$$k = 1$$

$$i = 6$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=6**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	1	2	0	0	0	0	0

$$k = 2$$

$$i = 7$$

- Nema poklapanja. Pošto je  $k > 0$  znamo da je pre toga bilo poklapanja. Promenljiva  $k$  dobija vrednost funkcije na poziciji  $k-1$ , dakle,  $k$  dobija vrednost 1.

$$k = 1$$

$$i = 7$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	1	2	0	0	0	0	0



$$k = 1$$

$$i = 7$$

- Nema poklapanja. Pošto je  $k > 0$  znamo da je pre toga bilo poklapanja. Promenljiva  $k$  dobija vrednost funkcije na poziciji  $k-1$ , dakle,  $k$  dobija vrednost 0.

$$k = 0$$

$$i = 7$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablona	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	1	2	0	0	0	0	0

$$k = 0$$

$$i = 7$$

- Nema poklapanja. Pošto je  $k=0$  zaključujemo da ne postoji prefiks jednak sufiksu.
- Povećavamo  $i$ .

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	1	2	0	0	0	0	0

$$k = 0$$

$$i = 8$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=8**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	1	2	0	1	0	0	0

$$k = 1$$

$$i = 9$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=9**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	1	2	0	1	2	0	0

$$k = 2$$

$$i = 10$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=10**. Promenljiva **i** se povećava.

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11  
Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	1	2	0	1	2	3	0



$$k = 3$$

$$i = 11$$

- Nema poklapanja. Pošto je  $k > 0$  znamo da je pre toga bilo poklapanja. Promenljiva  $k$  dobija vrednost funkcije na poziciji  $k-1$ , dakle,  $k$  dobija vrednost 2.

$$k = 2$$

$$i = 11$$

# KMP – primer funkcije neuspeha

0 1 2 3 4 5 6 7 8 9 10 11

Šablon: a a a b b a a b a a a a

Indeks	0	1	2	3	4	5	6	7	8	9	10	11
Karakter šablonu	a	a	a	b	b	a	a	b	a	a	a	a
Vrednost funkcije	0	1	2	0	0	1	2	0	1	2	3	3

$$k = 2$$

$$i = 11$$

- Poklapanje. Promenljiva **k** se uvećava. Upisuje se na poziciju **i=11**. Promenljiva **i** se povećava.
- Novo **i** dostiže dužinu kolekcije. Kraj postupka.

# KMP – primer algoritma

Šablon:      0 1 2 3 4  
                a b b a b

Indeks	0	1	2	3	4
Karakter šablonu	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b c c b c a b b a a b b a b  
      | | {  
      a b b a b

- Na indeksu **i=2** (posle **k=2** poklapanja) nailazimo na razliku. Pošto znamo da je prethodno bilo poklapanja (**k>0**) iščitavamo vrednost iz tabele na poziciji **k-1** i dodeljujemo je promenljivoj **k**.

$k = 0$   
 $i = 2$

# KMP – primer algoritma

Indeks	0	1	2	3	4
Karakter šablonu	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b c c b c a b b a a b b a b  
| {  
a b b a b

- Na indeksu **i=3** (posle **k=1** poklapanja) nailazimo na razliku. Pošto znamo da je prethodno bilo poklapanja (**k>0**) iščitavamo vrednost iz tabele na poziciji **k-1** i dodeljujemo je promenljivoj **k**.

$$k = 0$$

$$i = 3$$

# KMP – primer algoritma

Indeks	0	1	2	3	4
Karakter šablonu	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b c c b c a b b a a b b a b  
| | {  
a b b a b

- Na indeksu **i=5** (posle **k=2** poklapanja) nailazimo na razliku. Pošto znamo da je prethodno bilo poklapanja (**k>0**) iščitavamo vrednost iz tabele na poziciji **k-1** i dodeljujemo je promenljivoj **k**.

$$k = 0$$

$$i = 5$$

# KMP – primer algoritma

Indeks	0	1	2	3	4
Karakter šablonu	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b c c b c a b b a a b b a b  
          ⋮  
          a b b a b

- Na indeksu **i=5** (posle **k=0** poklapanja) nailazimo na razliku. Pošto znamo da prethodno nije bilo poklapanja, **i** uvećavamo za 1.

$$k = 0$$

i = 6

# KMP – primer algoritma

Indeks	0	1	2	3	4
Karakter šablonu	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

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

Tekst: a b a a b c c b c a b b a a b b a b

۷

a b b a b

- Na indeksu **i=6** (posle **k=0** poklapanja) nailazimo na razliku. Pošto znamo da prethodno nije bilo poklapanja, **i** uvećavamo za 1.

$$k = 0$$

i = 7

# KMP – primer algoritma

Indeks	0	1	2	3	4
Karakter šablona	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

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

Tekst: a b a a b c c b c a b b a a b b a b

⋮

a b b a b

- Na indeksu **i=7** (posle **k=0** poklapanja) nailazimo na razliku. Pošto znamo da prethodno nije bilo poklapanja, **i** uvećavamo za 1.

$k = 0$

$i = 8$

# KMP – primer algoritma

Indeks	0	1	2	3	4
Karakter šablona	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

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

Tekst: a b a a b c c b c a b b a a b b a b

{

a b b a b

- Na indeksu **i=8** (posle **k=0** poklapanja) nailazimo na razliku. Pošto znamo da prethodno nije bilo poklapanja, **i** uvećavamo za 1.

$k = 0$

$i = 9$

# KMP – primer algoritma

Indeks	0	1	2	3	4
Karakter šablonu	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b c c b c a b b a a b b a b  
| | | |  $\delta$   
a b b a b

- Na indeksu **i=13** (posle **k=4** poklapanja) nailazimo na razliku. Pošto znamo da je prethodno bilo poklapanja (**k>0**) iščitavamo vrednost iz tabele na poziciji **k-1** i dodeljujemo je promenljivoj **k**.

$$k = 1$$

$$i = 13$$

# KMP – primer algoritma

Indeks	0	1	2	3	4
Karakter šablonu	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b c c b c a b b a a b b a b  
|  $\xi$   
a b b a b

- Iz tabele znamo da se prvi karakter šablonu i karakter na indeksu 12 teksta poklapaju, pa nema potrebe da to ponovo proveravamo.
- Na indeksu **i=13** nailazimo na razliku. Pošto znamo da je prethodno bilo poklapanja (**k>0**) iščitavamo vrednost iz tabele na poziciji **k-1** i dodeljujemo je promenljivoj **k**.

$k = 0$   
 $i = 13$

# KMP – primer algoritma

Indeks	0	1	2	3	4
Karakter šablonu	a	b	b	a	b
Vrednost funkcije	0	0	0	1	2

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Tekst: a b a a b c c b c a b b a a b b a b  
| | | | |  
a b b a b

- Uspevamo da poklopimo svaki karakter šablonu sa tekstrom.
- Kraj postupka.

# Zadatak 1

- Implementirajte nalaženje podstringa grubom silom (brute force).

# Zadatak 2

- Implementirajte nalaženje podstringa Boyer-Moore metodom.

# Zadatak 3

- Implementirajte nalaženje podstringa KMP (Knuth-Morris-Pratt) metodom.