

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

# Delo z nizi

Matevž Jekovec

Univerza v Ljubljani  
Fakulteta za računalništvo in informatiko

Poletna šola 2014 — Programiranje v višji prestavi

Delo z nizi

Matevž  
Jekovec

Uvod

▶  $S$  — vhodni niz znakov ali besedilo (ang. *string*)

Indeksiranje  
besedila

▶  $N$  — dolžina niza (število znakov)

Iskanje z  
oknom

▶  $\Sigma$  — abeceda

Podobnost  
med nizi

▶  $\sigma$  — velikost abecede

Vaje in naloge

---

<sup>1</sup>Pozor: Dejanska velikost niza v pomnilniku je lahko drugačna (*null-delimited* nizi, različna kodiranje znakov)

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

- ▶  $S$  — vhodni niz znakov ali besedilo (ang. *string*)
- ▶  $N$  — dolžina niza (število znakov)
- ▶  $\Sigma$  — abeceda
- ▶  $\sigma$  — velikost abecede
- ▶ Primer:
  - ▶  $S = \text{"BANANA"}$
  - ▶  $N^1 = 6$
  - ▶  $\Sigma = \{B, A, N\}$
  - ▶  $\sigma = 3$

---

<sup>1</sup>Pozor: Dejanska velikost niza v pomnilniku je lahko drugačna (*null-delimited* nizi, različna kodiranje znakov)

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

## Oblike besedil:

- ▶ **Strukturirana:** Besedilo je sestavljeno iz več besed. Običajno tudi iščemo po besedah. (programska koda, angleščina, slovenščina)
- ▶ **Nestrukturirana:** Besedilo nima jasnih besed ali pa je veliko besed sestavljenk. (nemščina, človeški genom, glasba)

## Oblike besedil:

- ▶ Strukturirana: Besedilo je sestavljeno iz več besed. Običajno tudi iščemo po besedah. (programska koda, angleščina, slovenščina)
- ▶ Nestrukturirana: Besedilo nima jasnih besed ali pa je veliko besed sestavljenk. (nemščina, človeški genom, glasba)

## Problemi pri strukturiranih besedilih:

- ▶ Ali se beseda  $P$  (ang. *pattern*) nahaja v besedilu?
- ▶ Kje in kolikokrat se beseda ponovi?
- ▶ Ali se katera koli beseda začne na  $P$ ? Ali je  $P$  koren katere koli besede? Ali je  $P$  pripona katere koli besede?

## Oblike besedil:

- ▶ Strukturirana: Besedilo je sestavljeno iz več besed. Običajno tudi iščemo po besedah. (programska koda, angleščina, slovenščina)
- ▶ Nestrukturirana: Besedilo nima jasnih besed ali pa je veliko besed sestavljenk. (nemščina, človeški genom, glasba)

## Problemi pri strukturiranih besedilih:

- ▶ Ali se beseda  $P$  (ang. *pattern*) nahaja v besedilu?
- ▶ Kje in kolikokrat se beseda ponovi?
- ▶ Ali se katera koli beseda začne na  $P$ ? Ali je  $P$  koren katere koli besede? Ali je  $P$  pripona katere koli besede?

## Problemi pri nestrukturiranih besedilih:

- ▶ Ali se vzorec  $P$  pojavi v besedilu?
- ▶ Kje in kolikokrat se vzorec pojavi?

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

## Kako iščemo?

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

## Kako iščemo? Odvisno od scenarija!



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

## Kako iščemo? Odvisno od scenarija!

- ▶ Besedilo ves čas isto, vzorci se menjajo.
- ▶ Vzorec ves čas isti, besedilo se menja.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Kako iščemo? Odvisno od scenarija!

- ▶ Besedilo ves čas isto, vzorci se menjajo.
- ▶ Vzorec ves čas isti, besedilo se menja.

Ideja: V prvem primeru najprej indeksiramo besedilo, kar traja nekaj časa, vendar bo iskanje po kazalu (in ne neposredno po besedilu) potem bistveno hitreje. V drugem primeru porabimo nekaj časa, da spravimo vzorec v obliko, s katero bi hitreje iskali po besedilu.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Kako shraniti strukturirano besedilo tako, da bomo lahko hitro iskali po njem?

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Kako shraniti strukturirano besedilo tako, da bomo lahko hitro iskali po njem?

Ideja: Slovar, ki preslika iskan ključ na mesta pojavitev ključa v besedilu. Kaj točno uporabiti?

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Kako shraniti strukturirano besedilo tako, da bomo lahko hitro iskali po njem?

Ideja: Slovar, ki preslika iskan ključ na mesta pojavitev ključa v besedilu. Kaj točno uporabiti?

- ▶ Dvojiško iskalno drevo

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Kako shraniti strukturirano besedilo tako, da bomo lahko hitro iskali po njem?

Ideja: Slovar, ki preslika iskan ključ na mesta pojavitev ključa v besedilu. Kaj točno uporabiti?

- ▶ Dvojiško iskalno drevo
- ▶ Zgoščevalna tabela

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Kako shraniti strukturirano besedilo tako, da bomo lahko hitro iskali po njem?

Ideja: Slovar, ki preslika iskan ključ na mesta pojavitev ključa v besedilu. Kaj točno uporabiti?

- ▶ Dvojiško iskalno drevo
- ▶ Zgoščevalna tabela
- ▶ Številsko drevo

Delo z nizi

**Matevž  
Jekovec**

Uvod

**Indeksiranje  
besedila**

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Ideja: Iskalno drevo, vozlišča hranijo nize, manjši niz je tisti, ki je po abecedi (*leksikografsko*) pred drugim.

Delo z nizi

Matevž  
Jekovec

Uvod

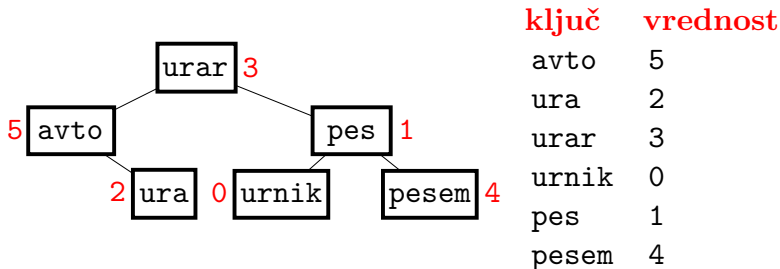
Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Ideja: Iskalno drevo, vozlišča hranijo nize, manjši niz je tisti, ki je po abecedi (*leksikografsko*) pred drugim.



Delo z nizi

Matevž  
Jekovec

Uvod

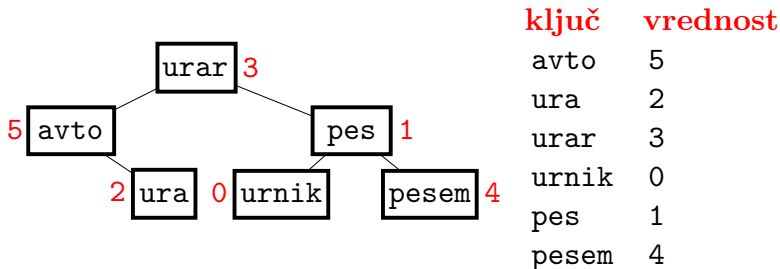
Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Ideja: Iskalno drevo, vozlišča hranijo nize, manjši niz je tisti, ki je po abecedi (*leksikografsko*) pred drugim.



Pozor: Če imamo nize že zložene v tabelo po abecedi, potem namesto iskanja po drevesu iščemo kar po tabeli — bisekcija!

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Zgoščeno vrednost izračunamo za vsako besedo v vhodnem besedilu. Nato vstavimo besedo kot običajni element (npr. število) v zgoščevalno tabelo.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Zgoščeno vrednost izračunamo za vsako besedo v vhodnem besedilu. Nato vstavimo besedo kot običajni element (npr. število) v zgoščevalno tabelo.

Rabin-Karpova zgoščevalna funkcija:

$$H = c_1 a^{k-1} + c_2 a^{k-2} + \dots + c_k a^0$$

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Zgoščeno vrednost izračunamo za vsako besedo v vhodnem besedilu. Nato vstavimo besedo kot običajni element (npr. število) v zgoščevalno tabelo.

Rabin-Karpova zgoščevalna funkcija:

$$H = c_1 a^{k-1} + c_2 a^{k-2} + \dots + c_k a^0$$

Primer zgoščene vrednosti za  $a = 2$ : URAR, dolžina je  $k = 4$   
U je 20. črka, R je 17. črka, A pa 0. črka po abecedi:

$$20 \cdot 2^3 + 17 \cdot 2^2 + 0 \cdot 2^1 + 17 \cdot 2^0 = 160 + 68 + 17 = 245$$

Pozor: Če želimo hraniti več pojavitev niza, potrebujemo *multimap*.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Vzemimo velikost polja  $M = 8$ .

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedilaIskanje z  
oknomPodobnost  
med nizi

Vaje in naloge

Vzemimo velikost polja  $M = 8$ .

$$H(\text{"avto"}) = 136 \quad \text{mod } 8 = 0$$

$$H(\text{"ura"}) = 114 \quad \text{mod } 8 = 2$$

$$H(\text{"urar"}) = 245 \quad \text{mod } 8 = 5$$

$$H(\text{"urnik"}) = 534 \quad \text{mod } 8 = 6$$

$$H(\text{"pes"}) = 86 \quad \text{mod } 8 = 0$$

$$H(\text{"pesem"}) = 364 \quad \text{mod } 8 = 0$$



Delo z nizi

Matevž  
Jekovec

Uvod

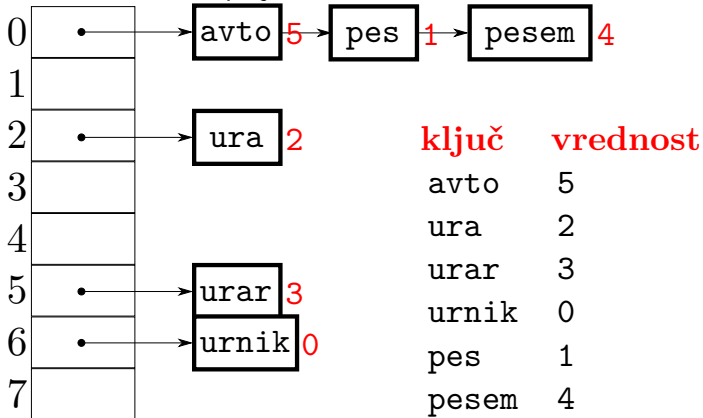
Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Vzemimo velikost polja  $M = 8$ .



Delo z nizi

Matevž  
Jekovec

Problem iskalnih dreves in zgoščevalne tabele: Ni mogoče iskati po začetkih besed (npr. *search-as-you-type*).

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

# Številsko drevo (ang. *trie*)

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

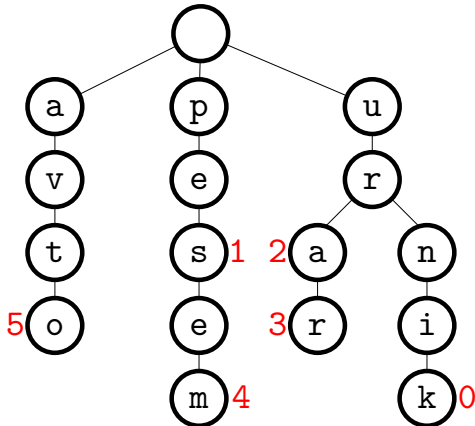
Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Problem iskalnih dreves in zgoščevalne tabele: Ni mogoče iskati po začetkih besed (npr. *search-as-you-type*).

Rešitev: Številsko drevo hrani besede od korena navzdol, razbite po črkah.



**ključ**      **vrednost**

avto      5

ura      2

urar      3

urnik      0

pes      1

pesem      4

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Izboljšava: Namesto svojega vozlišča za vsako črko, vozlišča z enim naslednikom združimo.

Delo z nizi

Matevž  
Jekovec

Uvod

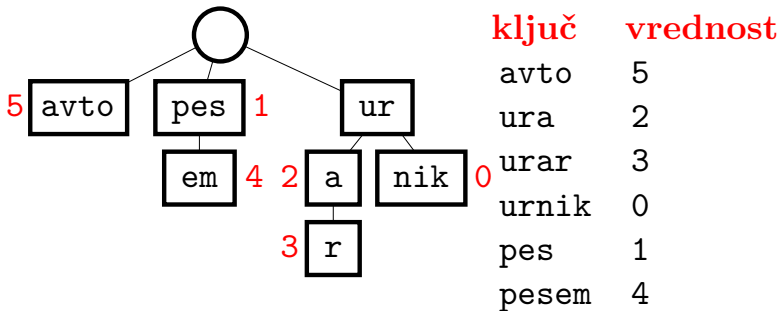
Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Izboljšava: Namesto svojega vozlišča za vsako črko, vozlišča z enim naslednikom združimo.



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

## Kako predstaviti vozlišče?

```
1 class TrieNode {  
2     string str;  
3     void *value;  
4     ?     children;  
5 };
```

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

## Kako predstaviti vozlišče?

```
1 class TrieNode {  
2     string str;  
3     void *value;  
4     ?     children;  
5 };
```

Do otrok dostopamo prek:

- ▶ povezanega seznama,

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

## Kako predstaviti vozlišče?

```
1 class TrieNode {  
2     string str;  
3     void *value;  
4     ?     children;  
5 };
```

Do otrok dostopamo prek:

- ▶ povezanega seznama,
- ▶ polje kazalcev v velikosti abecede  $\sigma$ ,



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

## Kako predstaviti vozlišče?

```
1 class TrieNode {  
2     string str;  
3     void *value;  
4     ?     children;  
5 };
```

Do otrok dostopamo prek:

- ▶ povezanega seznama,
- ▶ polje kazalcev v velikosti abecede  $\sigma$ ,
- ▶ trojiškega iskalnega drevesa,

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedilaIskanje z  
oknomPodobnost  
med nizi

Vaje in naloge

Vozlišče ima kazalec na prvega otroka in na sorojenca.

```
1 class TrieNode {
2     string    str;
3     void      *value;
4     TrieNode *child;
5     TrieNode *sibling;
6 };
```

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

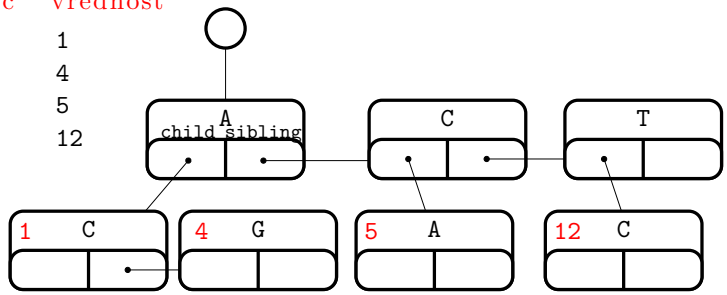
Vaje in naloge

Vozlišče ima kazalec na prvega otroka in na sorojenca.

```
1 class TrieNode {  
2     string    str;  
3     void      *value;  
4     TrieNode *child;  
5     TrieNode *sibling;  
6 };
```

ključ vrednost

AC	1
AG	4
CA	5
TC	12



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Vsako vozlišče vsebuje polje  $\sigma$  kazalcev na otroke. Do otroka dostopamo neposredno z indeksom njegove črke.

```
1 class TrieNode {
2     string    str;
3     void      *value;
4     TrieNode children[ALPHABET];
5 };
```

Delo z nizi

Matevž  
Jekovec

Vsako vozlišče vsebuje polje  $\sigma$  kazalcev na otroke. Do otroka dostopamo neposredno z indeksom njegove črke.

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

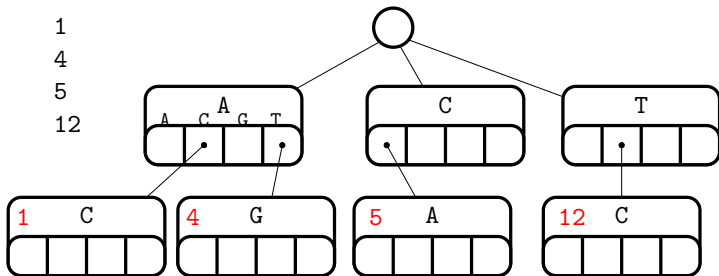
Vaje in naloge

```

1 class TrieNode {
2     string    str;
3     void      *value;
4     TrieNode children[ALPHABET];
5 };
    
```

ključ vrednost

AC	1
AG	4
CA	5
TC	12



Delo z nizi

Matevž  
Jekovec

Trojiško iskalno drevo (ang. *Ternary Search Trie*) je kompromis obeh prejšnjih metod: hranimo le kazalce na otroke, ki obstajajo, iskanje pa poteka z bisekcijo.

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

```

1 class TrieNode {
2     string    str;
3     void     *value;
4     TrieNode *eq;
5     TrieNode *ls;
6     TrieNode *gt;
7 };
  
```

Delo z nizi

Matevž  
Jekovec

Trojiško iskalno drevo (ang. *Ternary Search Trie*) je kompromis obeh prejšnjih metod: hranimo le kazalce na otroke, ki obstajajo, iskanje pa poteka z bisekcijo.

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

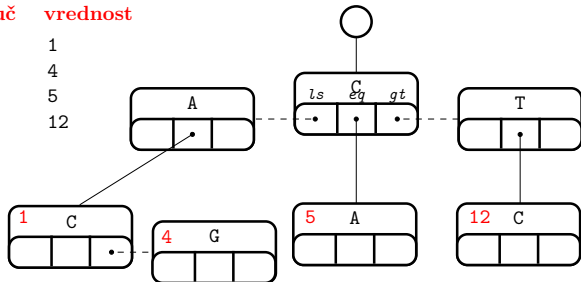
Vaje in naloge

```

1 class TrieNode {
2     string    str;
3     void      *value;
4     TrieNode *eq;
5     TrieNode *ls;
6     TrieNode *gt;
7 };
    
```

**ključ**    **vrednost**

AC	1
AG	4
CA	5
TC	12



Delo z nizi

Matevž  
Jekovec

Številsko drevo omogoča iskanje po začetkih besed. Kaj pa po korenu ali priponi?

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Številsko drevo omogoča iskanje po začetkih besed. Kaj pa po korenu ali priponi?

Priponsko drevo je stisnjeno številsko drevo, ki hrani vse predpone besedila (ali besed).

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Številsko drevo omogoča iskanje po začetkih besed. Kaj pa po korenu ali priponi?

Priponsko drevo je stisnjeno številsko drevo, ki hrani vse predpone besedila (ali besed).

1 2 3 4 5 6 7 8 9 10 11 12  
S=ABRAKADABRA\$

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Številsko drevo omogoča iskanje po začetkih besed. Kaj pa po korenu ali priponi?

Priponsko drevo je stisnjeno številsko drevo, ki hrani vse predpone besedila (ali besed).

1 2 3 4 5 6 7 8 9 10 11 12  
S=ABRAKADABRA\$

1	\$	12
2	A\$	11
3	ABRA\$	8
4	ABRAKADABRA\$	1
5	ADABRA\$	6
6	AKADABRA\$	4
7	BRA\$	9
8	BRAKADABRA\$	2
9	DABRA\$	7
10	KADABRA	5
11	RA\$	10
12	RAKADABRA\$	13

# Priponsko drevo (ang. *Suffix tree*)

Delo z nizi

Matevž  
Jekovec

Številsko drevo omogoča iskanje po začetkih besed. Kaj pa po korenu ali priponi?

Priponsko drevo je stisnjeno številsko drevo, ki hrani vse predpone besedila (ali besed).

Uvod

Indeksiranje  
besedila

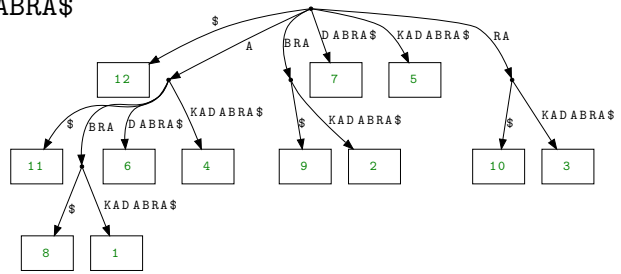
Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

1 2 3 4 5 6 7 8 9 10 11 12  
S=ABRAKADABRA\$

- 1 \$ 12
- 2 A\$ 11
- 3 ABRA\$ 8
- 4 ABRAKADABRA\$ 1
- 5 ADABRA\$ 6
- 6 AKADABRA\$ 4
- 7 BRA\$ 9
- 8 BRAKADABRA\$ 2
- 9 DABRA\$ 7
- 10 KADABRA 5
- 11 RA\$ 10
- 12 RAKADABRA\$ 13



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO BANANE.

P: BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO BANANE.

P: BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO BANANE.

P: BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO BANANE.

P: BANANA



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO BANANE.

P: BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO BANANE.

P: BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO BANANE.

P: BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO **BANANE**.

P: **BANANA**

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO **BANANE**.

P: **BANANA**

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO **BANANE**.

P: **BANANA**

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO **BANANE**.

P: **BANANA**

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO **BANANE**.

P: **BANANA**

Vaja: Koliko časa potrebujemo?



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

S: NA DREVESU RASTEJO **BANANE.**

P: **BANANA**

Vaja: Koliko časa potrebujemo?

Vsaj  $N$  primerjav znakov, če se vedno že prvi znak okna ne ujema in  $NM/2$ , če je vsak  $M$ -ti znak besedila zgrešitev. npr.

$S = \text{AAABAAABAAABAAAB}$  in  $P = \text{AAAA}$ .

Delo z nizi

Matevž  
Jekovec

Imejmo  $S = AAABAAABAAABAAAB$  in  $P = AAAA$  iz prejšnjega primera.

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Delo z nizi

Matevž  
Jekovec

Imejmo  $S = AAABAAABAAABAAAB$  in  $P = AAAA$  iz prejšnjega primera.

Uvod

Indeksiranje  
besedila

Prva zgrešitev se pojavi, ko je okno na poziciji  $i = 0$  in znaku  $j = 3$ , če štejemo od 0 dalje.

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Delo z nizi

Matevž  
Jekovec

Imejmo  $S = AAABAAABAAABAAAB$  in  $P = AAAA$  iz prejšnjega primera.

Uvod

Prva zgrešitev se pojavi, ko je okno na poziciji  $i = 0$  in znaku  $j = 3$ , če štejemo od 0 dalje.

Indeksiranje  
besedila

Iskanje z  
oknom

Ideja: Namesto, da premaknemo okno za 1 znak naprej, ga premaknemo za 4 znake naprej, saj znaka B sploh nimamo v vzorcu!

Podobnost  
med nizi

Vaje in naloge

Delo z nizi

Matevž  
Jekovec

Imejmo  $S = AAABAAABAAABAAAB$  in  $P = AAAA$  iz prejšnjega primera.

Uvod

Prva zgrešitev se pojavi, ko je okno na poziciji  $i = 0$  in znaku  $j = 3$ , če štejemo od 0 dalje.

Indeksiranje  
besedila

Iskanje z  
oknom

Ideja: Namesto, da premaknemo okno za 1 znak naprej, ga premaknemo za 4 znake naprej, saj znaka B sploh nimamo v vzorcu!

Podobnost  
med nizi

Vaje in naloge

Algotem: Zgradimo tabelo, ki hrani število znakov, ki jih lahko preskočimo glede na trenutni znak besedila.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Imejmo  $S = AAABAAABAAABAAAB$  in  $P = AAAA$  iz prejšnjega primera.

Prva zgrešitev se pojavi, ko je okno na poziciji  $i = 0$  in znaku  $j = 3$ , če štejemo od 0 dalje.

Ideja: Namesto, da premaknemo okno za 1 znak naprej, ga premaknemo za 4 znake naprej, saj znaka B sploh nimamo v vzorcu!

Algotem: Zgradimo tabelo, ki hrani število znakov, ki jih lahko preskočimo glede na trenutni znak besedila. Primera:

- ▶ Vzorec AAAA: vedno lahko preskočimo kar vse 4 znake, če se pojavi kateri koli znak  $\neq A$  na vhodu.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Imejmo  $S = \text{AAABAAABAAABAAAB}$  in  $P = \text{AAAA}$  iz prejšnjega primera.

Prva zgrešitev se pojavi, ko je okno na poziciji  $i = 0$  in znaku  $j = 3$ , če štejemo od 0 dalje.

Ideja: Namesto, da premaknemo okno za 1 znak naprej, ga premaknemo za 4 znake naprej, saj znaka B sploh nimamo v vzorcu!

Algotem: Zgradimo tabelo, ki hrani število znakov, ki jih lahko preskočimo glede na trenutni znak besedila. Primera:

- ▶ Vzorec AAAA: vedno lahko preskočimo kar vse 4 znake, če se pojavi kateri koli znak  $\neq A$  na vhodu.
- ▶ Vzorec ABAB: če se pojavi  $B$  na vhodu in pričakujemo  $A$ , potem preskočimo le en znak. Isto velja obratno. Če je na vhodu kateri koli drug znak, preskočimo 4 znake.

Delo z nizi

Matevž  
Jekovec

Preskočna tabela vsebuje indekse najbolj desnih pojavitev vseh možnih znakov abecede  $\Sigma$ .

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

$P = \text{ B A N A N A }$



Delo z nizi

Matevž  
Jekovec

Preskočna tabela vsebuje indekse najbolj desnih pojavitev vseh možnih znakov abecede  $\Sigma$ .

Uvod

Indeksiranje  
besedila

$$P = \begin{array}{cccccc} B & A & N & A & N & A \\ 0 & 1 & 2 & 3 & 4 & 5 \end{array}$$

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Delo z nizi

Matevž  
Jekovec

Preskočna tabela vsebuje indekse najbolj desnih pojavitev vseh možnih znakov abecede  $\Sigma$ .

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

	$P =$	B	A	N	A	N	A
$\Sigma:$		0	1	2	3	4	5
	A						
	B						
	C						
	D						
	E						
	...						
	M						
	N						
	O						
	...						

Delo z nizi

 Matevž  
 Jekovec

Preskočna tabela vsebuje indekse najbolj desnih pojavitev vseh možnih znakov abecede  $\Sigma$ .

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

	$P =$	B	A	N	A	N	A
$\Sigma:$		0	1	2	3	4	5
A	-1	-1	1	1	3	3	5
B	-1	0	0	0	0	0	0
C	-1	-1	-1	-1	-1	-1	-1
D	-1	-1	-1	-1	-1	-1	-1
E	-1	-1	-1	-1	-1	-1	-1
...							
M	-1	-1	-1	-1	-1	-1	-1
N	-1	-1	-1	2	2	4	4
O	-1	-1	-1	-1	-1	-1	-1
...							



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Iskanje: Okno postavimo na začetek  $i = 0$  in preverjamo znake od desne proti levi  $j = |P| - 1$ :

- ▶ Če se znak  $c$  v besedilu ujema z znakom v oknu, nadaljujemo s preverjanjem naslednjega znaka v oknu in besedilu.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Iskanje: Okno postavimo na začetek  $i = 0$  in preverjamo znake od desne proti levi  $j = |P| - 1$ :

- ▶ Če se znak  $c$  v besedilu ujema z znakom v oknu, nadaljujemo s preverjanjem naslednjega znaka v oknu in besedilu.
- ▶ Če se  $c$  ne ujema in je  $\text{right}[c] = -1$ , potem premakni okno za  $|P|$  znakov naprej.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Iskanje: Okno postavimo na začetek  $i = 0$  in preverjamo znake od desne proti levi  $j = |P| - 1$ :

- ▶ Če se znak  $c$  v besedilu ujema z znakom v oknu, nadaljujemo s preverjanjem naslednjega znaka v oknu in besedilu.
- ▶ Če se  $c$  ne ujema in je  $\text{right}[c] = -1$ , potem premakni okno za  $|P|$  znakov naprej.
- ▶ Če se  $c$  ne ujema in  $\text{right}[c] \neq -1$ , potem premaknemo okno za  $j - \text{right}[c]$  znakov. V primeru, da bi bila vrednost negativna, okno premaknemo za 1 znak naprej.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Iskanje: Okno postavimo na začetek  $i = 0$  in preverjamo znake od desne proti levi  $j = |P| - 1$ :

- ▶ Če se znak  $c$  v besedilu ujema z znakom v oknu, nadaljujemo s preverjanjem naslednjega znaka v oknu in besedilu.
- ▶ Če se  $c$  ne ujema in je  $\text{right}[c] = -1$ , potem premakni okno za  $|P|$  znakov naprej.
- ▶ Če se  $c$  ne ujema in  $\text{right}[c] \neq -1$ , potem premaknemo okno za  $j - \text{right}[c]$  znakov. V primeru, da bi bila vrednost negativna, okno premaknemo za 1 znak naprej.



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Iskanje: Okno postavimo na začetek  $i = 0$  in preverjamo znake od desne proti levi  $j = |P| - 1$ :

- ▶ Če se znak  $c$  v besedilu ujema z znakom v oknu, nadaljujemo s preverjanjem naslednjega znaka v oknu in besedilu.
- ▶ Če se  $c$  ne ujema in je  $\text{right}[c] = -1$ , potem premakni okno za  $|P|$  znakov naprej.
- ▶ Če se  $c$  ne ujema in  $\text{right}[c] \neq -1$ , potem premaknemo okno za  $j - \text{right}[c]$  znakov. V primeru, da bi bila vrednost negativna, okno premaknemo za 1 znak naprej.

S: NA DREVESU RASTEJO BANANE.

P: BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Iskanje: Okno postavimo na začetek  $i = 0$  in preverjamo znake od desne proti levi  $j = |P| - 1$ :

- ▶ Če se znak  $c$  v besedilu ujema z znakom v oknu, nadaljujemo s preverjanjem naslednjega znaka v oknu in besedilu.
- ▶ Če se  $c$  ne ujema in je  $\text{right}[c] = -1$ , potem premakni okno za  $|P|$  znakov naprej.
- ▶ Če se  $c$  ne ujema in  $\text{right}[c] \neq -1$ , potem premaknemo okno za  $j - \text{right}[c]$  znakov. V primeru, da bi bila vrednost negativna, okno premaknemo za 1 znak naprej.

S: NA DREVESU RASTEJO BANANE.

P: BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Iskanje: Okno postavimo na začetek  $i = 0$  in preverjamo znake od desne proti levi  $j = |P| - 1$ :

- ▶ Če se znak  $c$  v besedilu ujema z znakom v oknu, nadaljujemo s preverjanjem naslednjega znaka v oknu in besedilu.
- ▶ Če se  $c$  ne ujema in je  $\text{right}[c] = -1$ , potem premakni okno za  $|P|$  znakov naprej.
- ▶ Če se  $c$  ne ujema in  $\text{right}[c] \neq -1$ , potem premaknemo okno za  $j - \text{right}[c]$  znakov. V primeru, da bi bila vrednost negativna, okno premaknemo za 1 znak naprej.

S: NA DREVESU RASTEJO BANANE.

P:

BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Iskanje: Okno postavimo na začetek  $i = 0$  in preverjamo znake od desne proti levi  $j = |P| - 1$ :

- ▶ Če se znak  $c$  v besedilu ujema z znakom v oknu, nadaljujemo s preverjanjem naslednjega znaka v oknu in besedilu.
- ▶ Če se  $c$  ne ujema in je  $\text{right}[c] = -1$ , potem premakni okno za  $|P|$  znakov naprej.
- ▶ Če se  $c$  ne ujema in  $\text{right}[c] \neq -1$ , potem premaknemo okno za  $j - \text{right}[c]$  znakov. V primeru, da bi bila vrednost negativna, okno premaknemo za 1 znak naprej.

S: NA DREVESU RASTEJO BANANE.

P:

BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Iskanje: Okno postavimo na začetek  $i = 0$  in preverjamo znake od desne proti levi  $j = |P| - 1$ :

- ▶ Če se znak  $c$  v besedilu ujema z znakom v oknu, nadaljujemo s preverjanjem naslednjega znaka v oknu in besedilu.
- ▶ Če se  $c$  ne ujema in je  $\text{right}[c] = -1$ , potem premakni okno za  $|P|$  znakov naprej.
- ▶ Če se  $c$  ne ujema in  $\text{right}[c] \neq -1$ , potem premaknemo okno za  $j - \text{right}[c]$  znakov. V primeru, da bi bila vrednost negativna, okno premaknemo za 1 znak naprej.

S: NA DREVESU RASTEJO BANANE.

P: BANANA

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Uporabimo znanje iz zgoščevalnih tabel:

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Uporabimo znanje iz zgoščevalnih tabel:

1. Najprej izračunamo zgoščeno vrednost vzorca  $H_P$ .

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Uporabimo znanje iz zgoščevalnih tabel:

1. Najprej izračunamo zgoščeno vrednost vzorca  $H_P$ .
2. Postavimo okno na začetek  $i = 0$  in izračunamo zgoščeno vrednost znakov pod oknom  $H_S$ .



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Uporabimo znanje iz zgoščevalnih tabel:

1. Najprej izračunamo zgoščeno vrednost vzorca  $H_P$ .
2. Postavimo okno na začetek  $i = 0$  in izračunamo zgoščeno vrednost znakov pod oknom  $H_S$ .
3. Če  $H_S = H_P$ , preverimo vse znake pod oknom, da se prepričamo o enakosti, in vrnemo mesto pojavitve.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Uporabimo znanje iz zgoščevalnih tabel:

1. Najprej izračunamo zgoščeno vrednost vzorca  $H_P$ .
2. Postavimo okno na začetek  $i = 0$  in izračunamo zgoščeno vrednost znakov pod oknom  $H_S$ .
3. Če  $H_S = H_P$ , preverimo vse znake pod oknom, da se prepričamo o enakosti, in vrnemo mesto pojavitve.
4. Če  $H_S \neq H_P$ , se premaknemo za en znak naprej, izračunamo novo zgoščeno vrednost znakov pod oknom  $H_S$  in gremo na korak 3.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Ideja: Če imamo pametno zgoščevalno funkcijo, nam ni treba prebrati ponovno vseh znakov pod oknom, ampak le dodamo zgoščeno vrednost novega znaka in odstranimo zgoščeno vrednost najstarejšega znaka. Tako pohitrimo iskanje.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Ideja: Če imamo pametno zgoščevalno funkcijo, nam ni treba prebrati ponovno vseh znakov pod oknom, ampak le dodamo zgoščeno vrednost novega znaka in odstranimo zgoščeno vrednost najstarejšega znaka. Tako pohitrimo iskanje.  
Robin-Karpova zgoščevalna funkcija:

$$H = c_1 a^{|P|-1} + c_2 a^{|P|-2} + \dots + c_{|P|} a^0$$

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Ideja: Če imamo pametno zgoščevalno funkcijo, nam ni treba prebrati ponovno vseh znakov pod oknom, ampak le dodamo zgoščeno vrednost novega znaka in odstranimo zgoščeno vrednost najstarejšega znaka. Tako pohitrimo iskanje. Robin-Karpova zgoščevalna funkcija:

$$H = c_1 a^{|P|-1} + c_2 a^{|P|-2} + \dots + c_{|P|} a^0$$

Ko pride nov znak  $c'$ , odstranimo najstarejši znak  $c_{old}$  in izračunamo nov

$$H' = H \cdot a - c_{old} \cdot a^{|P|} + c'$$

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Levenshteinova razdalja ali *urejevalna razdalja*  $\delta$  med dvema nizoma  $S_1$  in  $S_2$  je najmanjše število vstavljanj, brisanj ali zamenjav posameznih znakov, da spremenimo  $S_1$  v  $S_2$ .

Delo z nizi

Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

Matevž  
Jekovec

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

Izračun razdalje se naredi z Levenshteinovo matriko:



Delo z nizi

Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Izračun razdalje se naredi z Levenshteinovo matriko:

B A N A N A

Delo z nizi

 Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Izračun razdalje se naredi z Levenshteinovo matriko:

	B	A	N	A	N	A
0	1	2	3	4	5	6
A	1					
N	2					
A	3					
N	4					
A	5					
S	6					

Delo z nizi

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

Izračun razdalje se naredi z Levenshteinovo matriko:

	B	A	N	A	N	A	
0	1	2	3	4	5	6	
A	1	1	1	2	3	4	5
N	2						
A	3						
N	4						
A	5						
S	6						

Delo z nizi

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

Izračun razdalje se naredi z Levenshteinovo matriko:

		B	A	N	A	N	A
	0	1	2	3	4	5	6
A	1	1	1	2	3	4	5
N	2	2	2	1	2	3	4
A	3						
N	4						
A	5						
S	6						

Delo z nizi

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

Izračun razdalje se naredi z Levenshteinovo matriko:

		B	A	N	A	N	A
	0	1	2	3	4	5	6
A	1	1	1	2	3	4	5
N	2	2	2	1	2	3	4
A	3	3	2	2	1	2	3
N	4						
A	5						
S	6						

Delo z nizi

Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Izračun razdalje se naredi z Levenshteinovo matriko:

		B	A	N	A	N	A
	0	1	2	3	4	5	6
A	1	1	1	2	3	4	5
N	2	2	2	1	2	3	4
A	3	3	2	2	1	2	3
N	4	4	3	2	2	1	2
A	5						
S	6						

Delo z nizi

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

Izračun razdalje se naredi z Levenshteinovo matriko:

		B	A	N	A	N	A
	0	1	2	3	4	5	6
A	1	1	1	2	3	4	5
N	2	2	2	1	2	3	4
A	3	3	2	2	1	2	3
N	4	4	3	2	2	1	2
A	5	5	4	3	2	2	1
S	6						

Delo z nizi

 Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Izračun razdalje se naredi z Levenshteinovo matriko:

		B	A	N	A	N	A
	0	1	2	3	4	5	6
A	1	1	1	2	3	4	5
N	2	2	2	1	2	3	4
A	3	3	2	2	1	2	3
N	4	4	3	2	2	1	2
A	5	5	4	3	2	2	1
S	6	6	5	4	3	3	2



Delo z nizi

 Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Izračun razdalje se naredi z Levenshteinovo matriko:

		B	A	N	A	N	A
	0	1	2	3	4	5	6
A	1	1	1	2	3	4	5
N	2	2	2	1	2	3	4
A	3	3	2	2	1	2	3
N	4	4	3	2	2	1	2
A	5	5	4	3	2	2	1
S	6	6	5	4	3	3	2

Levenshteinova razdalja se nahaja spodnjem desnem kotu  $\delta(|S_1|, |S_2|)$ .

Delo z nizi

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Formula za izračun razdalje med nizi  $S_1[0..i]$  in  $S_2[0..j]$ :

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + 1 \\ \delta(i, j-1) + 1 \\ \delta(i-1, j-1) + 1 & \text{če } S_1[i] \neq S_2[j] \\ \delta(i-1, j-1) & \text{če } S_1[i] = S_2[j] \end{cases}$$

Izračun razdalje se naredi z Levenshteinovo matriko:

		B	A	N	A	N	A
	0	1	2	3	4	5	6
A	1	1	1	2	3	4	5
N	2	2	2	1	2	3	4
A	3	3	2	2	1	2	3
N	4	4	3	2	2	1	2
A	5	5	4	3	2	2	1
S	6	6	5	4	3	3	2

Levenshteinova razdalja se nahaja spodnjem desnem kotu  $\delta(|S_1|, |S_2|)$ . Pot nam pove, katere ukaze je potrebno izvesti.

Delo z nizi

Matevž  
Jekovec

Za izračun najdaljšega skupnega podniza lahko uporabimo priponsko drevo obeh nizov in pogledamo najgloblje skupno vozlišče.

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Delo z nizi

Matevž  
Jekovec

Za izračun najdaljšega skupnega podniza lahko uporabimo priponsko drevo obeh nizov in pogledamo najgloblje skupno vozlišče.

Uvod

Lahko pa uporabimo matriko, podobno Levenshteinovi, le da spremenimo pogoj:

Indeksiranje  
besedila

Iskanje z  
oknom

$$LCSuff(S_1[1..i], S_2[1..j]) =$$

Podobnost  
med nizi

$$\begin{cases} LCSuff(S_1[1..i-1], S_2[1..j-1]) + 1 & \text{če } S_1[i] = S_2[j] \\ 0 & \text{sicer} \end{cases}$$

Vaje in naloge

Delo z nizi

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Za izračun najdaljšega skupnega podniza lahko uporabimo priponsko drevo obeh nizov in pogledamo najgloblje skupno vozlišče.

Lahko pa uporabimo matriko, podobno Levenshteinovi, le da spremenimo pogoj:

$$\begin{aligned}
 &LCSuff(S_1[1..i], S_2[1..j]) = \\
 &\begin{cases} LCSuff(S_1[1..i-1], S_2[1..j-1]) + 1 & \text{če } S_1[i] = S_2[j] \\ 0 & \text{sicer} \end{cases}
 \end{aligned}$$

		B	A	N	A	N	A
	0	0	0	0	0	0	0
A	0	0	1	0	1	0	1
N	0	0	0	2	0	2	0
A	0	0	1	0	3	0	3
N	0	0	0	2	0	4	0
A	0	0	0	0	3	0	5
S	0	0	0	0	0	0	0

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Najdaljše skupno podzaporedje (ang. *longest common subsequence*) je podobno najdaljšemu skupnemu podnizu, le da se lahko med posameznimi znaki, ki so prisotni v obeh nizih, vrinjeni tudi tuji znaki.

Delo z nizi

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Najdaljše skupno podzaporedje (ang. *longest common subsequence*) je podobno najdaljšemu skupnemu podnizu, le da se lahko med posameznimi znaki, ki so prisotni v obeh nizih, vrinjeni tudi tuji znaki.

$$LCS(S_1[1..i], S_2[1..j]) =$$

$$\begin{cases} LCS(S_1[1..i-1], S_2[1..j-1]) + 1 & \text{če } S_1[i] = S_2[j] \\ \max(LCS(S_1[1..i], S_2[1..j-1]), LCS(S_1[1..i-1], S_2[1..j])) & \\ \text{če } S_1[i] \neq S_2[j] \end{cases}$$

Delo z nizi

 Matevž  
 Jekovec

Uvod

 Indeksiranje  
 besedila

 Iskanje z  
 oknom

 Podobnost  
 med nizi

Vaje in naloge

Najdaljše skupno podzaporedje (ang. *longest common subsequence*) je podobno najdaljšemu skupnemu podnizu, le da se lahko med posameznimi znaki, ki so prisotni v obeh nizih, vrinjeni tudi tuji znaki.

$$LCS(S_1[1..i], S_2[1..j]) =$$

$$\begin{cases} LCS(S_1[1..i-1], S_2[1..j-1]) + 1 & \text{če } S_1[i] = S_2[j] \\ \max(LCS(S_1[1..i], S_2[1..j-1]), LCS(S_1[1..i-1], S_2[1..j])) & \text{če } S_1[i] \neq S_2[j] \end{cases}$$

		B	A	N	A	N	A
	0	0	0	0	0	0	0
A	0	0	1	1	1	1	1
N	0	0	1	2	2	2	2
A	0	0	1	2	3	3	3
N	0	0	1	2	3	4	4
A	0	0	1	2	3	4	5
S	0	0	1	2	3	4	5



Delo z nizi

**Matevž  
Jekovec**

Burkhard-Kellerjevo drevo se uporablja pri črkovalnikih in je uporabno za iskanje podobnih besed.

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

**Podobnost  
med nizi**

Vaje in naloge

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

Burkhard-Kellerjevo drevo se uporablja pri črkovalnikih in je uporabno za iskanje podobnih besed.

Gradnja: Začnemo s poljubnim pojmom. Nato dodamo novo geslo tako, da izračunamo urejevalno razdaljo  $D$  med trenutnim vozliščem in novim geslom. Če povezava v smeri razdalje ne obstaja, novo geslo pripnemo obstoječemu vozlišču. Sicer sledimo povezavi in izračunamo urejevalno razdaljo med novim vozliščem in našim geslom.

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedilaIskanje z  
oknomPodobnost  
med nizi

Vaje in naloge

Burkhard-Kellerjevo drevo se uporablja pri črkovalnikih in je uporabno za iskanje podobnih besed.

Gradnja: Začnemo s poljubnim pojmom. Nato dodamo novo geslo tako, da izračunamo urejevalno razdaljo  $D$  med trenutnim vozliščem in novim geslom. Če povezava v smeri razdalje ne obstaja, novo geslo pripnemo obstoječemu vozlišču. Sicer sledimo povezavi in izračunamo urejevalno razdaljo med novim vozliščem in našim geslom.

Iskanje najbolj podobnih besed z urejevalno razdaljo  $\delta$ :  
Začnemo pri korenu in izračunamo urejevalno razdaljo  $D$  med iskanim geslom in trenutnim vozliščem. Če je  $D \leq \delta$ , potem izpišemo trenutno vozlišče. Preiskujemo tiste otroke, ki imajo razdaljo  $\geq D - \delta$  ali  $\leq D + \delta$ .

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

gesla

ura

urar

urnik

avto

pes

pesem

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

gesla

ura

ura

urar

urnik

avto

pes

pesem

Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

gesla

ura

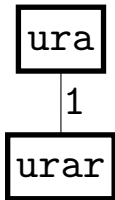
urar

urnik

avto

pes

pesem



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

gesla

ura

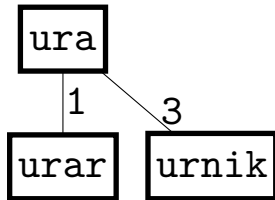
urar

urnik

avto

pes

pesem



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

gesla

ura

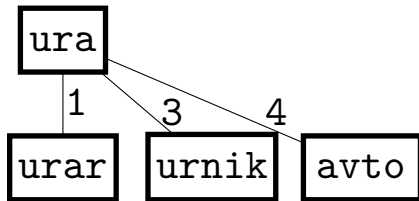
urar

urnik

avto

pes

pesem





Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

gesla

ura

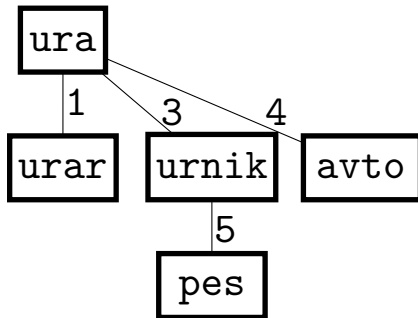
urar

urnik

avto

pes

pesem



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

gesla

ura

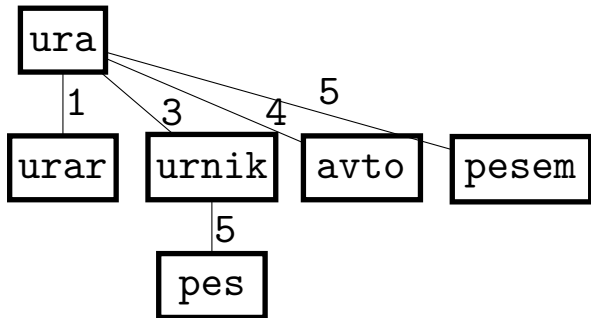
urar

urnik

avto

pes

pesem



Delo z nizi

Matevž  
Jekovec

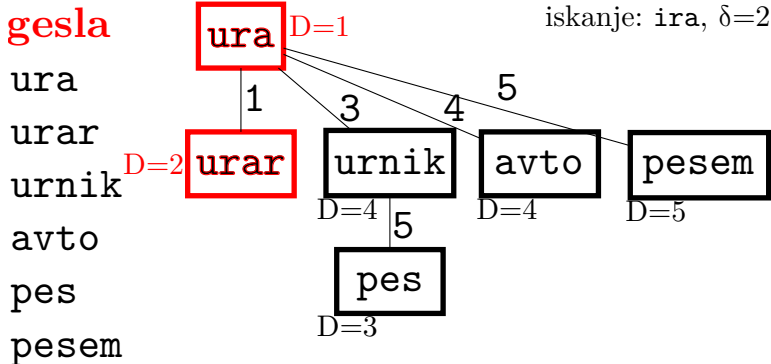
Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge



Delo z nizi

Matevž  
Jekovec

Uvod

Indeksiranje  
besedila

Iskanje z  
oknom

Podobnost  
med nizi

Vaje in naloge

## Vaje:

- ▶ S pomočjo urejanja (npr. quicksort) uredi vnešene nize po abecedi.
- ▶ Sprogramiraj številsko drevo, ki podpira operaciji vstavljanja in iskanja.
- ▶ Sprogramiraj BK-drevo, ki podpira operaciji vstavljanja in iskanja.
- ▶ Sprogramiraj Edit Distance, Longest Common Substring in Longest Common Subsequence s pomočjo dinamičnega programiranja.

## UVa naloge:

499 What's The Frequency, Kenneth?

454 Anagrams

164 String Computer - Edit distance

290 Palindroms  $\leftrightarrow$  smordnilaP

335 Processing MX Records

455 Periodic Strings