

Regularni izrazi

abeceda $\Sigma = \{a, b, c, \dots, z\}$ ali $\Sigma = \{0, 1\}$

besede: konini nizi simbolov $\Sigma^* = \{s_1 \dots s_n \mid s_i \in \Sigma \text{ za } i \leq n\}$
prazna beseda - označimo ϵ

$L \subseteq \Sigma^*$ jezik, vprašanje: ali je dana $w \in \Sigma$ element L ?

Primer:

L = množica nizov, ki predstavljajo veljaven datum

"2038-02-29" $\in L$?

"1-+(:" $\in L$

Regularni izrazi opisuje družino jezikov, regularni jeziki

↪ vzorci, ki opisujejo besede

Regularni jeziki:

\emptyset prazen jezik

$\{a\}$ za vsak $a \in \Sigma$, enojec

$L_1 \cup L_2$ če sta L_1 in L_2 regularna

$L_1 L_2 := \{w_1 w_2 \mid w_1 \in L_1, w_2 \in L_2\}$ stik jezikov L_1 in L_2

$L^* := \{w_1 w_2 \dots w_n \mid n \geq 0, w_i \in L \text{ za } i \leq n\}$ iteracija

$$w_1 = s_1 \dots s_m$$

$$w_2 = t_1 \dots t_n$$

$$w_1 w_2 = s_1 \dots s_m t_1 \dots t_n$$

$$\text{Primjer: } \{\{a\}\{\{b\}\} = \{\{ab\}\}$$

$$(\{\{a\}\} \cup \{\{b\}\}) \{\{c\}\} = \{\{a,b\}\} \{\{c\}\} = \{\{ac\}, \{bc\}\}$$

$$\{\{a,b\}\}^* = \{\emptyset, \{a,b\}, \{aa, ab, ba, bb\}, \{aaa, aab, \dots\}\}$$

$$\{\{c\}\}^* = \{\emptyset, \{c\}, \{cc\}, \{ccc\}, \{cccc\}, \dots\}$$

Regularni izraz

0

a

$L_1 + L_2$

$L_1 \cdot L_2$ ali $L_1 L_2$

L^*

1

Regularni jeziki

\emptyset

$\{\{a\}\}$

$L_1 \cup L_2$

$L_1 L_2$

L^*

$\{\emptyset\}$

Python

a

$L_1 | L_2$

$L_1 L_2$

L^*

$\{\emptyset\}$

Vježba:

$$0 + L = L$$

$$L_1 + L_2 = L_2 + L_1 \quad \dots$$

$$1 \cdot L = L$$

$$(L_1 + L_2) \cdot L_3 = L_1 \cdot L_3 + L_2 \cdot L_3$$

$$L^* = 1 + L \cdot L^*$$

$$x = 1 + L \cdot x$$

$$x = \frac{1}{1-L} = L^0 + L^1 + L^2 + L^3 + \dots$$

re N python

$$\Sigma = \{s_1, \dots, s_n\}$$

	kategorické znaky	$s_1 + s_2 + \dots + s_n$
R^*	nic ali vei ponovitev	R^*
R^+	ena ali več ponovitev	RR^*
$R^?$	nic ali ena ponovitev	$I+R$
$R\{7\}$	7 ponovitev	$RRRRRRR$
$R\{2,5\}$	2,3,4 ali 5 ponovitev	$RR+RRR+RRRR+RRRRR$

Ena ali dve zvezdici: $\{\ast\} \cup \{\ast\}\{\ast\}$

$$\ast + \ast\ast$$

python: $\ast | \ast\ast$
 $\backslash\ast | \backslash\ast\backslash\ast$ " \ast | \ast\ast "

" \ast | \ast\ast "



raw (\ nima posebnega pomena)

$[a,b,c,d,e]$ - - - - $a+b+c+d+e$

$[a-e]$ - - - - $a+b+c+d+e$

$[a-z]$ mala írkava

$[0-9]$ števka

$[a-z, A-Z]$ mala ali velika írkava

$([A-Z][a-z]^+)^2$ ImePríimok

$[A-Z][a-z]^+ [A-Z][a-z]^+$ Ime Príimok

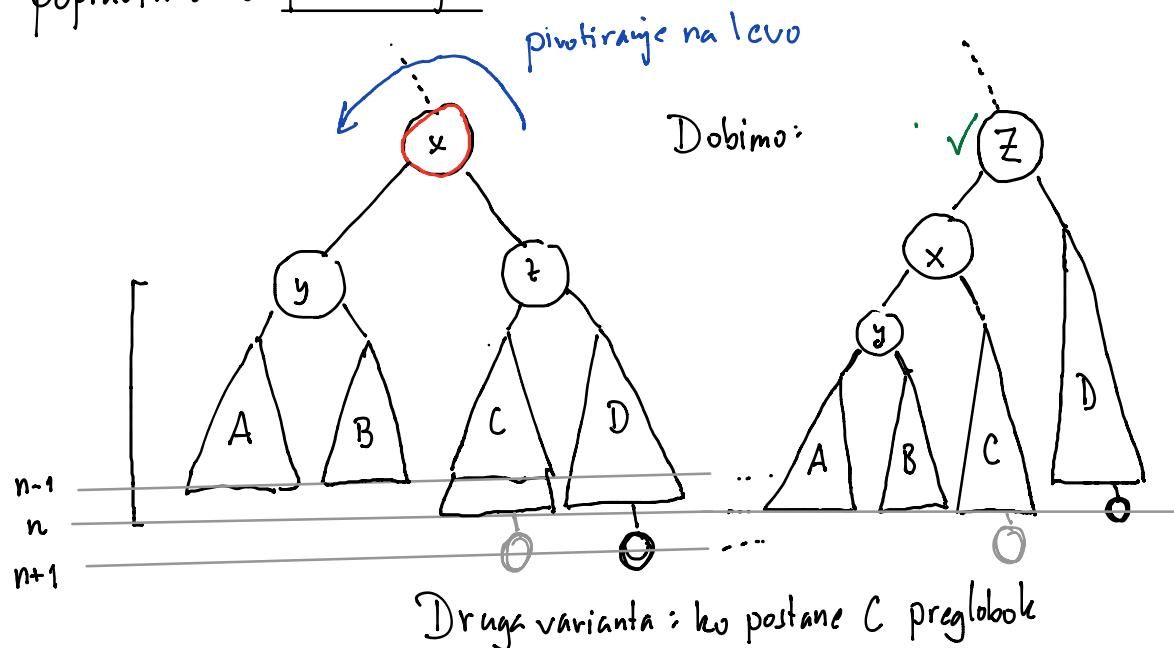
$[^{\wedge} \dots]$
↑
komplement

$[^{\wedge} a]$

vsi znaki ravn a

Od zadnjic: AVL drevesa

- Popravimo s pivotiranjem:



Druga varijanta:

