

Regularni izrazi

Σ osnovni simboli npr. $\Sigma = \{a, b, c, \dots, A, B, C, \dots, 0, \dots, 9, +, -, \times, \dots\}$
abeceda (končna) ☺

Σ^* množica vseh končnih zaporedij znakov (besede)

Zanimajo nas vzorci:

- telefonska številka
- datum
- email
- vejica, ki ji ne sledi presledek ali znak za novo vrsto

} opisuje množico besed

Jezik L je množica besed.

Osnovne operacije na jezikih:

- $\emptyset$ prazen jezik
- $\{a\}$ za $a \in \Sigma$
- $\{\epsilon\}$ kjer je ϵ prazna beseda (Python: "")
- unija: $L \cup M = \{w \mid w \in L \text{ ali } w \in M\}$

Primer: $\{a\} \cup \{c\} \cup \{m\} = \{a, c, m\}$

- stik: $L \cdot M = \{uv \mid u \in L \text{ in } v \in M\}$

Primer: $\{a, b\} \{c, d\} = \{ac, bc, ad, bd\}$

"
 $(\{a\} \cup \{b\}) (\{c\} \cup \{d\})$ $\nearrow$ prazna beseda
 $\epsilon aa = aa$

"
 $\{a, \epsilon\} \{aa\} = \{aaa, aa\}$

- iteracija ali ponavljanje

$$L^* = \{w_1 w_2 \dots w_n \mid w_1, \dots, w_n \in L\} \quad n \geq 0$$

Primeri:

$$\{a\}^* = \{\varepsilon, a, aa, aaa, aaaa, \dots\}$$

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

$$\{foo, bar\}^* = \{\varepsilon, foo, bar, foobar, foofoo, barfoo, barbar, foofoofoo, \dots\}$$

Regularni jezik je jezik, ki ga lahko dobimo z zgornjimi operacijami.

$$\Sigma^* = \{a, b, c, \dots, A, B, C, \dots, 0, \dots, 9\}^* \quad \text{vse besede}$$

Regularne jezike opisujemo z regularnimi izrazi

Regularni Jezik	Regularni Izraz	Python (regular expression, regex, rexp)
$\emptyset$	$\emptyset$	
$\{\varepsilon\}$	ε	" (prazen niz)
$\{a\}$	a	'a'
$L \cup M$	$P + Q$	$P Q$
$L \cdot M$	$P \cdot Q$ ali PQ	PQ
L^*	P^*	P^* (0 ali več ponovitev p)
$\{c_1, \dots, c_n\}$	$c_1 + c_2 + \dots + c_n$	$[c_1 c_2 \dots c_n]$
$\{A, B, C, \dots, Z\}$	$A + B + C + \dots + Z$	$[A-Z]$
$\{A, \dots, Z, a, \dots, z, 0, \dots, 9\}$	$A + \dots + Z + a + \dots + z + 0 + \dots + 9$	$[A-Za-z0-9]$
$L^+ = L \cdot L^*$	$P^+ = PP^*$	P^+ (ena ali več ponovitev)
$L = \{\varepsilon\} \cup L$	$P^? = \varepsilon + P$	$P^?$ (nič ali ena ponovitev)

Veljajo algebrski zakoni:

$$P + Q = Q + P$$

$$(P + Q) + R = P + (Q + R) \quad \emptyset + P = P$$

$$(P + Q) \cdot R = P \cdot R + Q \cdot R$$

$$P \cdot (Q + R) = P \cdot Q + P \cdot R \quad \varepsilon \cdot P = P \cdot \varepsilon = P$$

$$P^* = \varepsilon + PP^* = \varepsilon + P + P \cdot P + P \cdot P \cdot P + \dots$$

$$\hookrightarrow X = 1 + PX$$

$$(1 - P)X = 1$$

$$X = \frac{1}{1 - P} = 1 + P + P^2 + P^3 + P^4 + \dots$$

Regularni izrazi v Pythonu

ime & priimek:

Janez Novak
Ana Kovacic

$r "[A-Z][a-z]^+_[A-Z][a-z]^+ "$

$r "[A-Z][a-z]^+_[A-Z][a-z]^+ "$
2. skupina
1. skupina 3. skupina

Ponovitev dveh enakih steril, ločenih z vejico:

13, 15, 8, 7, 62, 62, 3, 4, 4, 4

$[1-9][0-9]^*|0$

0, 1, 10, 123, ...

$([1-9][0-9]^*|0),_*\backslash 1$
1. skupina

bolje $\backslash s^*$

$[0-9]^+$

0, 07, 13, 007, 80, ..., 00, ...