

# Regуларни изрази

## 1 Основни регуларни изрази

Регуларни изрази представљају механизам за препознавање одређених шаблона у тексту. Основни регуларни изрази су:

- $a$ : Ако је  $a$  карактер из азбуке, тада је  $a$  регуларни израз који препознаје само тај карактер.
- $r_1 r_2$ : Ако су  $r_1$  и  $r_2$  регуларни изрази, тада је  $r_1 r_2$  регуларни израз који препознаје низ где се  $r_2$  налази након  $r_1$ .
- $r_1 | r_2$ : Ако су  $r_1$  и  $r_2$  регуларни изрази, тада је  $r_1 | r_2$  регуларни израз који препознаје све што препознаје  $r_1$  или  $r_2$ .
- $r^*$ : За регуларни израз  $r$ , израз  $r^*$  препознаје 0 или више појављivanja израза  $r$ .

## 2 Проширенi регуларни изрази

Величина записа регуларног израза брзо расте како расте сложеност регуларног језика. Ради скраћивања записа и повећања читљивости, уводе се тзв. *проширенi регуларни изрази* који представљају краће записе неких основних регуларних израза.

Проширенi регуларни израз	Регуларни израз
$[a_1 a_2 \dots a_n]$	$a\_1   a\_2   \dots   a\_n$
$[^a_1 a_2 \dots a_n]$	Сви карактери из азбуке без $\{a_1, a_2, \dots, a_n\}$
$[a\_1 - a\_2]$	Дисјункција карактера од $a_1$ до $a_2$ из ASCII табеле
$\backslash w$	$[A-Za-z0-9\_]$
$\backslash W$	$[^A-Za-z0-9\_]$
$\backslash d$	$[0-9]$
$\backslash D$	$[^0-9]$
$\backslash s$	$[ \backslash t \backslash r \backslash n \backslash f ]$ (све бјeline)
$\backslash S$	$[^ \backslash t \backslash r \backslash n \backslash f ]$
$.$	$[^ \backslash n ]$ (све осим новог реда)
$a^+$	$aa^*$ (једно или више појављиванja)
$a^?$	$\varepsilon   a$ (нула или једно појављиванje)
$a^{\{n\}}$	$a^n$
$a^{\{n,m\}}$	$a^k, n \leq k \leq m$
$a^{\{n,\}}$	$a^k, n \leq k$

---

Karakteri sa specijalnim značenjem **u karakterskoj klasi**:

-, \, ], ^

Karakteri sa specijalnim značenjem **van karakterske klase**:

, +, ?, |, ., \, (, ), [, ...

Za oduzimanje specijalnog značenja koristi se \ (backslash).

### 3 Gramzivost i lenjo izračunavanje

**Gramzivost:** Poklapanje izraza vrši se u poslednjem mogućem trenutku. Gramzivi operatori su: \*, +, ?, { }

Gramzivost se uklanja dodatnim operatorom ? ukoliko je to podržano. Na primer: a+? je negramzivo poklapanje.

**Lenjo izračunavanje:** Ako se u disjunkciji prepozna prvi disjunkt, drugi se više ne razmatra.

### 4 Sidra

Sidra određuju položaj prepoznavanja regularnog izraza u tekstu, ne prepoznaju karaktere.

- ^ — početak teksta ili linije (u `multiline` režimu)
- \$ — kraj teksta ili linije (u `multiline` režimu)
- \b — granica reči (\w+ naspram \W)

### 5 Preduvidi i postuvidi

#### 5.1 Pozitivni preduvid

$r_1(? = r_2)$  izrazi  $r_1$  iza kojih стоји  $r_2$

$(? = r_1) r_2$  izrazi  $r_2$  koji почињу izrazom  $r_1$

#### 5.2 Negativni preduvid

$r_1(?!r_2)$  izrazi  $r_1$  iza kojih se ne nalazi  $r_2$

$(?!r_1) r_2$  izrazi  $r_2$  koji ne почињу  $r_1$

### 5.3 Pozitivni postuvid

$(? <= r_1) r_2$  izrazi  $r_2$  ispred kojih stoji  $r_1$

$r_1 (? <= r_2)$  izrazi  $r_1$  koji se završavaju izrazom  $r_2$

### 5.4 Negativni postuvid

$(? <!r_1) r_2$  izrazi  $r_2$  ispred kojih *ne* stoji  $r_1$

$r_1 (? <!r_2)$  izrazi  $r_1$  koji se *ne* završavaju izrazom  $r_2$

## 6 Grupisanje i referenciranje (Backreferences)

Grupisanje se vrši stavljanjem dijela izraza u zagrade  $()$ . Grupama se pristupa pomoću reference  $\backslash k$ , gde je  $k$  redni broj grupe. Numeracija odgovara redosledu otvorenih zagrada.