

Метакарактери

| () [] { } . * + ? ^ \$ \

Основне операције

Нека су r_1 и r_2 регуларни изрази, а s_1 и s_2 редом ниске које ти регуларни изрази препознају.

$r_1 r_2$
Дописивање (конкатенација). Препознаје ниску $s_1 s_2$ добијену дописивањем ниски s_1 и s_2 .

$r_1 | r_2$
Унија (алтернација). Препознаје сваку од ниски s_1, s_2 .

Џокери

Општи џокер

• џокер, мења произвољан карактер осим знака за нови ред.

Карактерске класе

Нека су $c_1, c_2, \dots, c_n$ произвољни карактери. Под карактером увек подразумевамо Unicode-карактер.

$[c_1 c_2 \dots c_n]$
џокер који мења један, било који од карактера $c_1, c_2, \dots, c_n$ (карактерска класа).

$[\hat{c}_1 c_2 \dots c_n]$
џокер који мења један карактер који није ниједан од карактера $c_1, c_2, \dots, c_n$ (негација карактерске класе).

$[c_1 - c_n]$
џокер који мења један, било који карактер уређене азбуке Unicode између карактера c_1 и c_n , укључујући и њих (опсег карактера).

$[\hat{c}_1 - c_n]$
џокер који мења један, било који карактер уређене азбуке Unicode, разликује се од карактера c_1, c_n и од сваког карактера између њих (негација опсега карактера).

Квантификатори (понављање)

Нека је r регуларни израз, а s произвољна ниска коју r препознаје. Нека су m и n произвољни природни бројеви, $0 \leq m \leq n$.

$r?$ препознаје 0 или 1 појављивање ниске s (s или празну ниску).

r^* препознаје коначан низ (од 0 или више) дописаних ниски од којих се свака може препознати регуларним изразом r (ниске у низу могу бити једнаке, а и не морају). Специјално, ако r препознаје тачно једну ниску s , тада r^* препознаје 0 или више појављивања ниске s ($s, ss, sss, \dots$, укључујући празну ниску). Број појављивања је произвољан природан број (укључујући 0).

r^+ препознаје коначан низ (од 1 или више) дописаних ниски од којих се свака може препознати регуларним изразом r (ниске у низу могу бити једнаке, а и не морају). Специјално, ако r препознаје тачно једну ниску s , тада r^+ препознаје 1 или више појављивања ниске s ($s, ss, sss, \dots$). Број појављивања је произвољан природан број већи од 0.

$r\{n\}$ препознаје коначан низ дописаних n ниски од којих се свака може препознати регуларним изразом r (ниске у низу могу бити једнаке, а и не морају). Специјално, ако r препознаје тачно једну ниску s , тада $r\{n\}$ препознаје тачно n узастопних појављивања ниске s ($\underbrace{s \dots s}_n$).

$r\{m, n\}$ препознаје коначан низ од најмање m и највише n дописаних ниски од којих се свака може препознати регуларним изразом r (ниске у низу могу бити једнаке, а и не морају). Специјално, ако r препознаје тачно једну ниску s , тада $r\{m, n\}$ препознаје k узастопних појављивања ниске s ($\underbrace{s \dots s}_k$), при чему је $0 \leq m \leq k \leq n$.

Могуће је изоставити n и у том случају нема ограничења на број појављивања: исто значење имају $r\{0, \}$ и r^* , односно $r\{1, \}$ и r^+ , а истог су значења и изрази $r\{0, 1\}$ и $r?$.

Сидра (контекстни селектори)

XML Schema их не користи. Користе их Notepad++ и програмски језици (Python, JavaScript, PHP).

$\wedge$ почетак линије (Notepad++), односно почетак ниске (програмски језици).

$\$$ крај линије (Notepad++), односно крај ниске (програмски језици).

Врачара

„Скида чини” метакарактерима (губе специјално значење, $\backslash$ је обична тачка итд.) и „баца чине” на обичне карактере (те добијају специјално значење).

Белине

$\backslash t$ табулатор.

$\backslash r$ знак за нови ред (енгл. carriage return, враћање колица на почетак линије).

$\backslash n$ знак за нови ред (енгл. line feed, прелазак у следећи ред).

Windows користи пар $\backslash r \backslash n$ као крај линије. Unix-деривати (Linux, FreeBSD) користе $\backslash n$ као крај линије. Мас користи $\backslash r$ као крај линије.

Специјалне карактерске класе

$\backslash s$ један карактер који је нека белина (размак, табулатор, знаци за нови ред).

$\backslash S$ један карактер који није белина.

$\backslash d$ једна декадна цифра.

$\backslash D$ један карактер који није декадна цифра.

Приоритет и груписање

Обичне заграде служе за груписање и имају највећи приоритет. Приоритет опада следећим редоследом:

- карактерске класе ($[]$),
- квантификатори ($* + ?$),
- дописивање и џокер ($.$),
- сидра ($\wedge \$$),
- унија или алтернација ($|$).