

Główne klasy znaków

Literały, znaki dosłowne Znaki mające taką samą wartość jak w piśmie, z którego pochodzą

Metaznaki Znaki lub ich kombinacje opisujące literały

Znaki ucieczki Zamieniają metaznak na literał (zob. niżej)

Metaznaki

`.` Dowolny znak (zob. niżej)

`a|b` Alternatywa (a lub b)

`[^-]` Klasy znaków (zob. niżej)

`*+?` Kwantyfikatory (zob. niżej)

`{}` Kwantyfikatory (zob. niżej)

`()` Grupa znaków (zob. niżej)

`^$` Granice tekstu (zob. niżej)

`\` Klasa znaków, znak ucieczki, znak biały (zob. niżej)

Znaki ucieczki

`\\` Ukośnik lewy

`\.` Kropka

`\[\]` Nawiasy kwadratowe

`\{ \}` Klamry

`\(\)` Nawiasy okrągłe

`\\` Rura

`\-` Łącznik (dywiz)

`\\^` Karetka

`\\$` Znak dolara

`*` Gwiazdka (asterisk)

`\\+` Plus

Znaki sterujące (białe)

`\\t` Tabulacja (pozioma; ASCII: 9)

`\\v` Tabulacja pionowa, dowolny pionowy znak biały

`\\V` Każdy znak za wyjątkiem pionowego znaku białego*

`\\n` Koniec akapitu (new line, line feed; ASCII: 10)**

`\\N` Każdy znak za wyjątkiem znaku końca akapitu*

`\\r` Powrót karetki (carriage return; ASCII: 13)

`\\R` Dowolna sekwencja znaków Unicode oznaczająca nowy akapit

`\\a` Dzwonek (alarm; ASCII: 7)

`\\0` Znak null (ASCII: 0)

*) Nie wszystkie implementacje.

***) W niektórych tekstach koniec akapitu (enter) to 2 znaki: `\\r\\n`.

Kropka – dowolny znak

Kropka oznacza każdy znak za wyjątkiem znaku (znaków) końca akapitu, a więc `\\n` lub `\\r\\n`. Może jednak oznaczać również te znaki po włączeniu odpowiedniej funkcji (np. `single line`, `dotall`) lub dodaniu flagi `/s` lub modyfikatora `(?s)`.

Klasy znaków

`[abc]` 1 znak: a albo b, albo c

`[a-c]` 1 znak z zakresu od a do c

`[^abc]` 1 znak różny od a oraz b, oraz c

`[^a-c]` 1 znak różny od każdego z zakresu od a do c

`\\d` Cyfry: `\\d = [0-9] = [0123456789]`

`\\D` Niecyfry: `\\D = [^0-9] = [0123456789]`

`\\w` Znak wyrazowy: `\\w = [a-zA-Z0-9_]`

`\\W` Znak niewyrazowy: `\\W = [^a-zA-Z0-9_]`

`\\s` Biały znak (spacja, enter, tabulacja i in.)

`\\S` Znak niebiały: `\\S = \\w + !, ?, @ i in.`

`\\L*` Małe litery

`\\U*` Wielkie litery

`\\E*` Wyłącza działanie `\\U` oraz `\\L`

`\\b` Granice wyrazu

`\\B` Granice niewyrazu

*) Działa tylko w oknie "Zamień na".



Kwantyfikatory			
Zachłanne	Leniwe	Zaborcze	Znak a...
a*	a*?	a*+	0 razy lub więcej
a+	a+?	a++	1 raz lub więcej
a?	a??	a?+	0 lub 1 raz
a{3}	a{3}?	a{3}+	3 razy
a{3,}	a{3,}?	a{3,}+	3 lub więcej razy
a{3,6}	a{3,6}?	a{3,6}+	od 3 do 6 razy

Kwantyfikatory zachłanne - zwracają maksymalne możliwe dopasowanie.
 Kwantyfikatory leniwe - zwracają minimalne możliwe dopasowanie.
 Kwantyfikatory zaborcze - zwracają tylko maksymalne dopasowanie lub żadne.

Grupy znaków	
Grupy przechwycone	
(abc)	Nienazwana grupa <i>abc</i>
(? ab cd)	Nienazwane grupy <i>ab</i> i <i>cd</i>
(?'x'ab)	Grupa <i>ab</i> o nazwie <i>x</i>
(?'<x>ab)	Grupa <i>ab</i> o nazwie <i>x</i>
(?'x'ab cd)	Grupy <i>ab</i> i <i>cd</i> o nazwie <i>x</i>
(?'<x>ab cd)	Grupy <i>ab</i> i <i>cd</i> o nazwie <i>x</i>
Grupy nieprzechwycone	
(?:abc)	Nienazwana grupa <i>abc</i>

Odwołania do przechwyconych grup znaków	
\1	Pierwsza grupa nienazwana
\g<1>	
\g{1}	
\g{-1}	Ostatnia grupa nienazwana (przed odwołaniem)
\g<x>	Grupa o nazwie <i>x</i>
\g<0>	Wszystkie wyszukania (także grupy nieprzechwycone)

Granice (kotwice)		
Początek	Koniec	
Granice niekonsumujące znaku (sprawdzają tylko obecność znaku granicznego)		
^	\$	Cały tekst lub akapit*
\b	\b	Wyraz
\B	\B	Niewyraz
\<	\>	Wyraz
[:<:]	[:>:]	Wyraz (zob. POSIX)
\A	\Z	Cały tekst
\Q	\E	Wyłączenie metaznaków
Granice konsumujące znaki (zwracają sam znak graniczny)		
(**)	\W	Wyraz
	\n lub \r\n	Akapit

*) Znaki ^, \$ oznaczają granice akapitu tylko w niektórych implementacjach lub przy użyciu flagi /m lub modyfikatora (?m) (multiline).

***) Konsumujące znaki końca wyrazu i akapitu wskazują też początek niepierwszego wyrazu i akapitu. Znakiem końca ostatniego wyrazu i akapitu są \$ oraz \Z.

Asercje (assertions)	
Przewidywania postępowe (lookaheads)	
a(?=b)	Zwróć <i>a</i> , jeśli PO nim jest <i>b</i>
a(?!b)	Zwróć <i>a</i> , jeśli PO nim NIE ma <i>b</i>
Przewidywania wsteczne (lookbehinds)	
(?<=b)a	Zwróć <i>a</i> , jeśli PRZED nim jest <i>b</i>
(?<!b)a, (?!=b)a	Zwróć <i>a</i> , jeśli PRZED nim NIE ma <i>b</i>
Wyrażenie warunkowe	
(?w)a b	Jeśli dopasowano <i>w</i> , znajdź <i>a</i> , jeśli nie, znajdź <i>b</i>
(?w)a	Jeśli dopasowano <i>w</i> , znajdź <i>a</i>
(?w) b	Jeśli NIE dopasowano <i>w</i> , znajdź <i>b</i>
(g)(?1)a	Warunek jako grupa nienazwana
(?'n')(?'n)a	Warunek jako grupa nazwana
(?(?<=w)a)	Warunek jako przewidywanie



Opcje (modifiers)

(?)	Pominięcie wielkości liter
(?-i)	Uwzględnienie wielkości liter
(?x)	Ignorowanie spacji i napisów po # (komentarzy)
(?-x)	Traktuj spacje i # jak literały
(?m)	^ i \$ wskazują początek/koniec akapitu
(?-m)	^ i \$ wskazują tylko początek/koniec tekstu
(?J)	Możliwe duplikaty nazw grup*

Zapis modyfikatorów

(?ism) ≡ (?i)(?s)(?m)	Kilka modyfikatorów
(?is) ≡ (?si)	Kolejność modyfikatorów
(?-im) ≡ (?-i)(?-m)	Negacja modyfikatorów
(?:te)st ≡ (?i)te(?-i)st	Zawężenie działania modyfikatorów

* Nie we wszystkich silnikach.

Opcje silnika PCRE (verbs)

(*UCP)	Poszerzenie klas znaków o znaki z całego Unicode*
(*LIMIT_MATCH=5)	Ograniczenie liczby dopasowań do 5
(*LIMIT_RECURSION=5)	Ograniczenie liczby dopasowywanych podwzorców

Ustawienia znaków i sekwencji oznaczających koniec akapitu:

(*CR)	Koniec akapitu to znak \r
(*LF)	Koniec akapitu to znak \n
(*CRLF)	Koniec akapitu to sekwencja \r\n
(*ANYCRLF)	Koniec akapitu to \r lub \n, lub \r\n

Opcje silnika PCRE (verbs) (cont)

(*ANY) Koniec akapitu do dowolny znak oznaczający w Unicode koniec akapitu**

) Klasy ld, lw, ls, lh, lv oznaczają domyślnie znaki z zakresu ASCII. Opcja (UCP) poszerza ich zakresy na całą tablicę Unicode.

** Znakami końca akapitu stają się: \r\n, \r, \n, \v, \f, U-0085, U-2028, U-2029.

Definicje wyszukiwań

(?(DEFINE) Zdefiniowanie wyszukiwania *abc* o nazwie *x* (?*x**abc*)

\g<*x*> Wywołanie wyszukiwania o nazwie *x*

\g<*x*>

Same definicje są ignorowane podczas wyszukiwania.

Macierz podstawiania

Zamiana wielu znaków w 2 krokach, np. *a, b, c* na, odpowiednio, *x, y, z*.

Krok:	Znajdź:	Zamień na:
1	([abc])	`\1'ax'by'cz`
2	`(.).*?\1(.).*?`	\2

Kody znaków

\x##	2-cyfrowy kod szesnastkowy (Unicode)
\x{#}	kod szesnastkowy (Unicode)
\o{#}	kod ósemkowy (Unicode)

Znaki POSIX

[:digit:]	Cyfry
[:xdigit:]	Cyfry szesnastkowe [0-9a-fA-F]
[:alpha:]	Litery
[:lower:]	Małe litery
[:upper:]	Wielkie litery
[:alnum:]	Litery i cyfry
[:ascii:]	Znaki z tablicy ASCII
[:punct:]	Nielitery i niecyfry
[:word:]	Znaki wyrazowe (litery, cyfry i podkreślnik)



Znaki POSIX (cont)

<code>[:graph:]</code>	Znaki widoczne (niebiałe)
<code>[:print:]</code>	Znaki widoczne (niebiałe) i spacje
<code>[:blank:]</code>	Spacja i tabulacja
<code>[:space:]</code>	Białe znaki
<code>[:cntrl:]</code>	Znaki kontrolne
<code>[:<:]</code>	Początek wyrazu
<code>[:>:]</code>	Koniec wyrazu

Grupy znaków Unicode

Główne grupy znaków Unicode

<code>\pC</code>	Znaki niewidoczne i nieużywane (kody bez znaków)
<code>\pL</code>	Litery
<code>\pN</code>	Liczby
<code>\pM</code>	Łączące znaki diakrytyczne
<code>\pP</code>	Znaki przestankowe (nielitery i niecyfry)
<code>\pS</code>	Symbole matematyczne, znaki walut, piktogramy, semigrafika itp.
<code>\pZ</code>	Białe znaki

Podgrupy znaków Unicode

<code>\p{Cc}</code>	Znaki kontrolne
<code>\p{Cf}</code>	Niewidoczne znaki formatowania
<code>\p{Cn}</code>	Kody bez przypisanych znaków
<code>\p{Co}</code>	Znaki do użytku prywatnego
<code>\p{Cs}</code>	Kody wykluczone z użytku (dot. kompatybilności z UTF-16)
<code>\p{Li}</code>	Litery małe z dowolnych języków
<code>\p{Lu}</code>	Litery wielkie z dowolnych języków
<code>\p{Lo}</code>	Litery bez odpowiednika wielkiego/małego
<code>\p{Lt}</code>	Dwuznaki o pojedynczym kodzie
<code>\p{Mc}</code>	Łączący znak diakrytyczny o szerokości litery
<code>\p{Me}</code>	Znak, w który wpisywany jest inny znak
<code>\p{Mn}</code>	Łączący znak diakrytyczny wpisywany w pole innej litery
<code>\p{Nd}</code>	Liczby dziesiętne

Grupy znaków Unicode (cont)

<code>\p{No}</code>	Cyfry w górnym i dolnym indeksie
<code>\p{Pc}</code>	Znak przestankowy łączący słowa (np. " _")
<code>\p{Pd}</code>	Dowolny myślnik lub łącznik
<code>\p{Ps}</code>	Dowolny nawias otwierający
<code>\p{Pe}</code>	Dowolny nawias zamykający
<code>\p{Pi}</code>	Dowolny znak otwierający cytata
<code>\p{Pf}</code>	Dowolny znak zamykający cytata
<code>\p{Po}</code>	Pozostałe widoczne znaki niealfanumeryczne
<code>\p{Sc}</code>	Symbole walut
<code>\p{Sk}</code>	Znaki modyfikujące odstępów (diakryty dopisywane)
<code>\p{Sm}</code>	Symbole matematyczne
<code>\p{So}</code>	Pozostałe symbole
<code>\p{Zl}</code>	Separator wiersza (U+2028)
<code>\p{Zp}</code>	Separator akapitu (U+2029)
<code>\p{Zs}</code>	Niewidoczny znak mający pewną szerokość