

正規表現早見表

Regex quick reference sheet

制作: 藍碌 (Airoku)



4.0 International

論理

*PHP, .NET のみ

パターン	意味
<code>expr ...</code>	または (<code>expr</code> でなければ ...)
<code>(?(cond)then else)*</code>	条件式 <code> else</code> は省略可 <code>cond</code> はアンカーかグループ番号/名前

量 化 子

パターン	意味
<code>expr*</code>	0 個以上の <code>expr</code>
<code>expr+</code>	1 個以上の <code>expr</code>
<code>expr?</code>	1 個以下の <code>expr</code>
<code>expr{n}</code>	ちょうど <code>n</code> 個の <code>expr</code>
<code>expr{n,}</code>	<code>n</code> 個以上の <code>expr</code>
<code>expr{n,m}</code>	<code>n</code> 個以上 <code>m</code> 個未満の <code>expr</code>
<code>ptrn?</code>	上記かつ、できるだけ少ない個数で

前方参照

*Go, Rust 非対応
†Python 非対応

パターン	意味
<code>\num*</code>	以前の番号付き部分式への一致
<code>\k<name>*†</code>	以前の名前付き部分式への一致

置 換

*JS 非対応
†JS, .NET のみ ‡.NET のみ

パターン	意味
<code>\$num</code>	番号付き部分式に置換
<code>\${name}*</code>	名前付き部分式に置換
<code>\$\$</code>	\$ に置換
<code>&†</code>	一致全体に置換
<code>\$~†</code>	一致より前の文字列全体に置換
<code>\$'†</code>	一致より後の文字列全体に置換
<code>\$+‡</code>	一致の最後のグループに置換
<code>\$_‡</code>	入力文字列全体に置換

1 文 字

*JS 非対応 †Go, Rust 非対応
‡PHP, Go 非対応 §Java は \Octal

パターン	意味	等価な例
<code>\n</code>	LF 改行	<code>\x0a</code>
<code>\r</code>	CR 改行	<code>\x0d</code>
<code>\t</code>	HT タブ	<code>\x09</code>
<code>\v</code>	VT 垂直タブ	<code>\x0b</code>
<code>\b</code>	BS バックスペース	<code>\x08</code>
<code>\e†</code>	ESC エスケープ	<code>\x1b</code>
<code>\f</code>	FF フォームフィード	<code>\x0c</code>
<code>\a</code>	BEL ベル	<code>\x07</code>
<code>\cchar†</code>	Ctrl+char に対応する制御文字	
<code>\char</code>	char の 1 文字	例: <code>\.</code> → <code>.NET</code>
<code>\octal§</code>	8 進数 (1-3 桁, 0-377)	<code>\123</code> → <code>S</code>
<code>\xhex</code>	16 進数 (1-2 桁, 0-ff)	<code>\x42</code> → <code>B</code>
<code>\uhex‡</code>	16 進数 (4 桁, UTF-16)	<code>\u3042</code> → <code>あ</code>

文字クラス

*JS, Python 非対応
†.NET のみ

パターン	意味	例
<code>.</code>	改行(\n)以外の 1 文字	<code>Asdf</code>
<code>[abc]</code>	a か b か c のどれか 1 文字	<code>bread</code>
<code>[^abc]</code>	a か b か c 以外の 1 文字	<code>caps</code>
<code>[a-z]</code>	Unicode で a から z の 1 文字	<code>LaTeX</code>
<code>[^A-Z]</code>	A から Z 以外の 1 文字	<code>SATySFi</code>
<code>\s</code>	空白文字 1 文字 [<code>\r\n\t\v\f</code>] と等価	
<code>\S</code>	\s はその補集合 (<code>[^\r\n\t\v\f]</code>)	
<code>\w</code>	英数字かアンダーバー 1 文字	
<code>\W</code>	<code>[A-Za-z0-9_]</code> と等価 \W はその補集合	
<code>\d</code>	数字 1 文字	
<code>\D</code>	<code>[0-9]</code> と等価 \D はその補集合	
<code>\p{cat}*</code>	Unicode プロパティで指定するどれか	
<code>\P{cat}*</code>	1 文字 (別表参照) \P はその補集合	
	5 以外の数字 (文字クラスの減算)	
<code>[0-9-[5]]†</code>	減算の引数 (0-9 と 5 にあたる部分) には補集合指定子 <code>~</code> を使える	

ア ン カ ー

*JS 非対応 †Python 非対応
‡Go, Rust 非対応

パターン	意味
<code>^</code>	文字列/行の先頭
<code>\$</code>	文字列/行の末尾か末尾の \n の直前
<code>\A*</code>	文字列の先頭
<code>\Z*†</code>	文字列の末尾か末尾の \n の直前
<code>\z*†</code>	文字列の末尾
<code>\G*††</code>	前回の一致の直後
<code>\b</code>	\w と \W の境界 \B ならそれ以外
<code>(?=expr)‡</code>	expr の直前
<code>(?!expr)‡</code>	expr 以外の直前
<code>(?<=expr)‡</code>	expr の直後
<code>(?<!expr)‡</code>	expr 以外の直後

グ ル ー プ

*Python 非対応 †JS 非対応 ‡Go 非対応
§Rust 非対応 ¶PHP, .NET のみ ||.NET のみ

パターン	意味
<code>(...)</code>	番号付きグループ
<code>(?<name>...)*</code>	名前付きグループ
<code>(?'name'...)*†</code>	name は <code>[A-Za-z_] \w*</code>
<code>(?<cl-op>...)*†</code>	op の最終一致から ... の直前までを
<code>(?'cl-op'...)*†</code>	cl に代入 op と cl はグループの名前
<code>(?:...)</code>	捕捉(代入)しないグループ
<code>(?i:...)*†</code>	...内では大文字小文字を区別しない
<code>(?m:...)*†</code>	...内では行ごと(複数行)モード
<code>(?s:...)*†</code>	...内では単一文字列モード
<code>(?n:...)*†</code>	...内の (...) を捕捉(代入)しない
<code>(?x:...)*††</code>	スペースを無視
<code>(?>...)*††§</code>	... を最初の一一致までのみ評価

そ の 他

JS 非対応 *Go 非対応 †Rust, Java 非対応
‡n は PHP, .NET のみ, x は Go 非対応

パターン	意味
<code>(?imnsx)‡</code>	以降のモードフラグを変更
<code>(?#comment)*†</code>	インラインコメント
<code>#comment*</code>	行コメント (x モードでのみ有効)



Unicode General Category Values
(UAX #44: Unicode Character Database)

<i>cat</i>	意味
Lu	大文字
Ll	小文字
Lt	タイトルケースのダイグラフ
Lm	修飾文字
Lo	その他の文字
L	Lu, Ll, Lt, Lm または Lo
Mn	ゼロ幅の結合文字
Mc	幅のある結合文字
Me	囲み結合文字
M	Mn, Mc または Me
Nd	十進法の数字
Nl	文字様の数字
No	その他の数字
N	Nd, Nl または No
Pc	タイ () 様の約物
Pd	ダッシュ (ー) 様の約物
Ps	開き側の約物
Pe	閉じ側の約物
Pi	始め引用符様の約物
Pf	終わり引用符様の約物
Po	その他の約物
P	Pc, Pd, Ps, Pe, Pi, Pf または Po
Sm	数学用記号
Sc	通貨記号
Sk	記号様の修飾文字
So	その他の記号
S	Sm, Sc, Sk または So
Zs	幅のあるスペース
Zl	U+2028 LINE SEPARATOR のみ
Zp	U+2029 PARAGRAPH SEPARATOR のみ
Z	Zs, Zl または Zp
Cc	C0・C1 制御文字
Cf	書式制御文字
Cs	サロゲート用コードポイント
Co	私用領域
Cn	非文字
C	“その他” Cc, Cf, Cs, Co または Cn

Unicode ブロック .NET のみ

始点	終点	<i>cat</i>
0000	007F	IsBasicLatin
0080	00FF	IsLatin-1Supplement
0100	017F	IsLatinExtended-A
0180	024F	IsLatinExtended-B
0250	02AF	IsIPAExtensions
02B0	02FF	IsSpacingModifierLetters
0300	036F	IsCombiningDiacriticalMarks
0370	03FF	IsGreek または IsGreekandCoptic
0400	04FF	IsCyrillic
0500	052F	IsCyrillicSupplement
0530	058F	IsArmenian
0590	05FF	IsHebrew
0600	06FF	IsArabic
0700	074F	IsSyriac
0780	07BF	IsThaana
0900	097F	IsDevanagari
0980	09FF	IsBengali
0A00	0A7F	IsGurmukhi
0A80	0AFF	IsGujarati
0B00	0B7F	IsOriya
0B80	0BFF	IsTamil
0C00	0C7F	IsTelugu
0C80	0CFF	IsKannada
0D00	0D7F	IsMalayalam
0D80	0DFF	IsSinhala
0E00	0E7F	IsThai
0E80	0EFF	IsLao
0F00	0FFF	IsTibetan
1000	109F	IsMyanmar
10A0	10FF	IsGeorgian
1100	11FF	IsHangulJamo
1200	137F	IsEthiopic
13A0	13FF	IsCherokee
1400	167F	IsUnifiedCanadianAboriginalSyllabics
1680	169F	IsOgham
16A0	16FF	IsRunic
1700	171F	IsTagalog
1720	173F	IsHanunoo
1740	175F	IsBuhid
1760	177F	IsTagbanwa
1780	17FF	IsKhmer
1800	18AF	IsMongolian
1900	194F	IsLimbu
1950	197F	IsTaiLe
19E0	19FF	IsKhmerSymbols
1D00	1D7F	IsPhoneticExtensions
1E00	1EFF	IsLatinExtendedAdditional
1F00	1FFF	IsGreekExtended
2000	206F	IsGeneralPunctuation
2070	209F	IsSuperscriptsandSubscripts
20A0	20CF	IsCurrencySymbols

本文書は .NET における正規表現の仕様をもとに作成し, regex101.com で検証した。

始点	終点	<i>cat</i>
20D0	20FF	IsCombiningDiacriticalMarksforSymbols または IsCombiningMarksforSymbols
2100	214F	IsLetterlikeSymbols
2150	218F	IsNumberForms
2190	21FF	IsArrows
2200	22FF	IsMathematicalOperators
2300	23FF	IsMiscellaneousTechnical
2400	243F	IsControlPictures
2440	245F	IsOpticalCharacterRecognition
2460	24FF	IsEnclosedAlphanumerics
2500	257F	IsBoxDrawing
2580	259F	IsBlockElements
25A0	25FF	IsGeometricShapes
2600	26FF	IsMiscellaneousSymbols
2700	27BF	IsDingbats
27C0	27EF	IsMiscellaneousMathematicalSymbols-A
27F0	27FF	IsSupplementalArrows-A
2800	28FF	IsBraillePatterns
2900	297F	IsSupplementalArrows-B
2980	29FF	IsMiscellaneousMathematicalSymbols-B
2A00	2AFF	IsSupplementalMathematicalOperators
2B00	2BFF	IsMiscellaneousSymbolsandArrows
2E80	2EFF	IsCJKRadicalsSupplement
2F00	2FDF	IsKangxiRadicals
2FF0	2FFF	IsIdeographicDescriptionCharacters
3000	303F	IsCJKSymbolsandPunctuation
3040	309F	IsHiragana
30A0	30FF	IsKatakana
3100	312F	IsBopomofo
3130	318F	IsHangulCompatibilityJamo
3190	319F	IsKanbun
31A0	31BF	IsBopomofoExtended
31F0	31FF	IsKatakanaPhoneticExtensions
3200	32FF	IsEnclosedCJKLettersandMonths
3300	33FF	IsCJKCompatibility
3400	4DBF	IsCJKUnifiedIdeographsExtensionA
4DC0	4DFF	IsYijingHexagramSymbols
4E00	9FFF	IsCJKUnifiedIdeographs
A000	A48F	IsYiSyllables
A490	A4CF	IsYiRadicals
AC00	D7AF	IsHangulSyllables
D800	DB7F	IsHighSurrogates
DB80	DBFF	IsHighPrivateUseSurrogates
DC00	DFFF	IsLowSurrogates
E000	F8FF	IsPrivateUse または IsPrivateUseArea
F900	FAFF	IsCJKCompatibilityIdeographs
FB00	FB4F	IsAlphabeticPresentationForms
FB50	FDFD	IsArabicPresentationForms-A
FE00	FE0F	IsVariationSelectors
FE20	FE2F	IsCombiningHalfMarks
FE30	FE4F	IsCJKCompatibilityForms
FE50	FE6F	IsSmallFormVariants
FE70	FEFF	IsArabicPresentationForms-B
FF00	FFEF	IsHalfwidthandFullwidthForms
FFF0	FFFF	IsSpecials