

# ■正規表現チートシート

(バックスラッシュ「\」は、お使いの環境に合わせて半角円マーク「¥」に読み替えて下さい)

表記	意味	補足説明
文字の <b>種類</b> に関するもの		
.	何でも。ただし改行以外	ドット。印刷だと見にくい、ピリオドのこと
[文字どれか]	[ ]内のどれかの文字	[A-Za-z]なら英大文字と小文字。[0-9A-F]なら16進数に使う文字。後に量の指定をつけなければ1個
[^これ以外]	[ ]内の文字以外どれも	[ ]内の先頭にあるとき、^は「否定」を意味する。ハット、キャレットなどと呼ぶ
\d	「0」～「9」の半角数字	[0-9]と同じ意味
\D	\d以外。つまり数字以外	[^0-9]と同じ。大文字は「否定」を意味する
\s	半角空白、タブ、改行、改ページなど	[ \t\n\r\f]に相当する。普通は <b>全角スペースは含まない</b>
\S	\s以外	\sの否定
\n	改行	\r\nと書く必要がある場合も。改行の文字コードが0Sによって異なり、WindowsはCR+LFのため
文字の <b>数</b> に関するもの		
*	何個でも	スター。とてもよく使う。 <b>0個（その文字がない）も該当するので注意</b>
+	1個以上	よく使う。すぐ上の「*」との違いがポイント。●●*や、●{1,*}と同じ意味
?	0個または1個	●{0,1}と同じ意味
{3}	3個ぴったり	●●●と同じ
{3,5}	3個～5個	●{3} ●{4} ●{5}と同じ。 <b>コンマの後に空白を入れないこと</b>
{3,}	3個以上いくつでも	●{3}●*と同じ。なお、「3個以下」と指示したいときは、●{0,3}とする
<b>位置</b> に関するもの（アンカーと呼ぶ。アンカーは、文字ではなくて位置にマッチする）		
^	行頭（複数行モードの場合）	^●●なら「行頭に●●がある場合」。この^は[ ]の外なので、行頭を指す
\$	行末（複数行モードの場合）	●●\$なら「行末に●●がある場合」
<b>その他</b>		
	または	パイプ。日本語キーボードなら右上隅にある。太郎 次郎 三郎、は[太次三]郎、と同じ意味。James Jack、はJa(mes ck)とも書ける。包含関係にあるものを並べる際は、長い方を前に書くほうが安全。社長秘書 社長、はOKだが、社長 社長秘書、はNG
(まとめる)	( )内をまとめ、かつ、後で引用できるよう番号を振って一時保存する	番号は外から中、左から右の順に振られる
\1、\2、\3…	( )で保存された内容	(.)りに\1る、なら「走りに走る」や「語りに語る」など。(.{3})\1と、なら、「ぽつんぽつん」や「ひらりひらりと」「どすんどすんと」など。\$1、\$2、\$3…と書く場合もある
\	直後の文字を、その文字本来の意味に戻す	エスケープ。特別な意味合いを解除する。\.なら「ピリオドそのもの」。\\なら「バックスラッシュそのもの」
/i	大文字小文字を区別しないモード	tokyo/i、ならば[tT][oO][kK][yY][oO]と同じ

※細かいことを言うと、環境によって、動作の仕方には、部分一致（該当するパターンが含まれていれば、前後に余分な文字があっても見つけ出してくれる）と、完全一致（文字列の最初から最後まで、ぴたりと一致する場合しか見つけ出してくれない）の2通りある。部分一致がデフォルトの状態、完全一致として動作させたいなら、行頭と行末の位置指定をつけ「**^ここに正規表現を書く\$**」とする。完全一致がデフォルトの状態、部分一致として動作させたいなら、前後にドットとスターを補って「**.\*ここに正規表現を書く.\***」のようにすればよい