

Regular Expressions

github.com/JKramerDD
Jeremy Kramer

[] - Square Brackets:

Indicate to include whatever symbol(s) are inside of the brackets:

[wW]	: The woodchuck named Walter drank water.
[0-9]	: He went jogging 13 times in the last 42 days.
[A-Z]	: A man named John Carter was a legend.
[a-z]	: A man named John Carter was a legend.

^ - Carat:

Indicate negation only when first in square brackets

[^A-Z]	: NOT A-Z
[^Ss]	: NOT S OR s
[^e^]	: NOT e OR ^

At beginning of line when used outside square brackets, indicates the first character in a line

^[A-Z]	: Palo Alto
^[^A]	: Everybody

| - Pipe:

Indicates an 'Or' operation

groundhog woodchuck	: Groundhog or woodchuck
a b c	: a or b or c

? - Question Mark:

Indicates the previous character is optional

[Cc]olou?r	: color, colour, Color, Colour
------------	--------------------------------

* - Star:

Matches 0 or more of the previous character

Oo*h	: Oh, Ooh, Oooh, ...
------	----------------------

+ - Plus:

Indicates 1 or more of the previous character

o+h	: oh, ooh, oooh, ooooh, ...
-----	-----------------------------

. - Period:

indicates any character

beg.n	: begin, begun, began, beg2n, begtn ...
-------	---

\$ - Dollar Sign:

Indicates to match previous character with the last character in the expression

\.\$: The end.
[a-z]\$: The fox ambushed the rabbit
.\$: The end?

\ - Backslash:

Used as an escape character to indicate the symbol within an expression rather than the Reg-Ex command

\.\$: The end.
\?	: Did you know? How did you find out?

Generating Accurate Reg-Ex:

Common Errors:

1) False positives (Type 1)

[tT]he	: theology, other
[cC]at	: Catastrophe, Cataclysm
[iI]	: igloo, ideology, mississippi

2) False negatives (Type 2)

the	: The
cat	: Cat

When generating regular expressions, be sure to go through a phase of checking the accuracy of your regular expression. There are various tools available online to assist you in checking your reg-ex strings, such as: www.regexpal.com