

GREP & Metacharacters 2015

in the InDesign CC 2015 Find/Change dialog

Description	Text	GREP	Description	Text	GREP	Description	GREP	Found (Change Field)	GREP
Tab	^t	^t	Any Double Quotation Mark	"	"	Beginning of Word	\<	Found Text	\$0
Forced Line Break	^n	^n	Any Single Quotation Mark	'	'	End of Word	\>	Found 1	\$1
End of Paragraph	^p	^r	Straight Double Quotation Mark	"	~"	Word Boundary	\b	Found 2	\$2
Any Line End Character (NEW!)		^R	Double Left Quotation Mark	^{"	~{"	Beginning of Paragraph	^	Found 3	\$3
Symbols	Text	GREP	Double Right Quotation Mark	^}"	~}"	End of Paragraph	\$	Found 4	\$4
Bullet Character	^8	~8	Straight Single Quotation Mark	^{'	~{'	Beginning of Story	^A	Found 5	\$5
Japanese Bullet	^5	~5	Single Left Quotation Mark	^[{'	~[{'	End of Story	^Z	Found 6	\$6
Carat Character	^	^	Single Right Quotation Mark	^}{'	~}{'	Repeat	GREP	Found 7	\$7
Backslash Character	^ \	^ \	Break Character	Text	GREP	Zero or One Time	?	Found 8	\$8
Copyright Symbol	^2	~2	Standard Carriage Return	^b	~b	Zero or More Times	*	Found 9	\$9
Ellipsis	^e	~e	Column Break	^M	~M	One or More Times	+	More undocumented GREP codes	
Tilde	~	^~	Frame Break	^R	~R	Zero or One Time (Shortest Match)	??	Letter	\p{letter}
Paragraph Symbol	^7	~7	Page Break	^P	~P	Zero or More Times (Shortest Match)	*?	Lowercase letter	\p{lowercase_letter}
Registered Trademark Symbol	^r	~r	Odd Page Break	^L	~L	One or More Times (Shortest Match)	+?	Uppercase letter	\p{uppercase_letter}
Section Symbol	^6	~6	Even Page Break	^E	~E	Match	GREP	Titlecase letter	\p{titlecase_letter}
Trademark Symbol	^d	~d	Discretionary Line Break	^k	~k	Marking Subexpression	()	Spacing modifier letter	\p{modifier_letter}
Open Parenthesis	(^(\	Variable	Text	GREP	Non-Marking Subexpression	(?.)	Non-case non-modifier letters	\p{letter_other}
Closed Parenthesis)	^)\	Any Variable	^v	~v	Character Set	[]	Marks	\p{mark}
Open Brace Character	{	^{	Running Header (Paragraph Style)	^Y	~Y	Or		Diacritical and tone marks	\p{non_spacing_mark}
Closed Brace Character	}	^}\	Running Header (Character Style)	^Z	~Z	Positive Lookbehind	(?<=)	Vowels SE Asian languages	\p{spacing_combining_mark}
Open Bracket Character	[^[Custom Text Variable	^u	~u	Negative Lookbehind	(?<!)	Circles, squares, keycaps	\p{enclosing_mark}
Closed Bracket Character]	^]\	Last Page Number Variable	^T	~T	Positive Lookahead	(?=)	Spaces, returns, 2028, 2029	\p{separator}
Markers	Text	GREP	Chapter Number Variable	^H	~H	Negative Lookahead	(?!)	All spaces except tab & return	\p{space_separator}
Any Page Number	^#	~#	Creation Date Variable	^O	~O	Lookbehind (NEW!)	\K	Line separator character 2028	\p{line_separator}
Current Page Number	^N	~N	Modification Date Variable	^O	~O	(Keep text found so far out of overall Regexp match)		Paragraph separator 2029	\p{paragraph_separator}
Next Page Number	^X	~X	Output Date Variable	^D	~D	Modifiers	GREP	Math and Wingdings	\p{symbol}
Previous Page Number	^V	~V	File Name Variable	^1	~1	Case-Insensitive On	(?i)	Math symbols	\p{math_symbol}
Section Marker	^x	~x	Metadata Caption	^J	~J	Case-Insensitive Off	(?i)	Currency symbols	\p{currency_symbol}
Anchored Object Marker	^a	~a	Other	Text	GREP	Multiline On or Multiline Off	(?m) or (?-m)	Modifier symbols	\p{modifier_symbol}
Footnote Reference Marker	^F	~F	Right Indent Tab	^y	~y	Single-Line On or Single-Line Off	(?s) or (?-s)	Wingdings, dingbats	\p{other_symbol}
Index Reference Marker	^I	~I	Indent to Here	^i	~i	Ignore Whitespace On	(?w)	Numbers	\p{number}
Hyphens and Dashes	Text	GREP	End Nested Style Here	^h	~h	Ignore Whitespace Off	(?w)	Digits 0 to 9	\p{decimal_digit_number}
Em Dash	^_	~_	Non-Joiner	^j	~j	Posix	GREP	Roman numerals	\p{letter_number}
En Dash	^=	~=	Clipboard Contents Formatted	^c	~c	Any alphanumeric character	[:alnum:]	Other numbers	\p{other_number}
Discretionary Hyphen	^-	~-	Clipboard Contents Unformatted	^C	~C	Any alphabetic character	[:alpha:]	Punctuation	\p{punctuation}
Nonbreaking Hyphen	^~	~~	Wildcards	Text	GREP	Only digits 0 to 9	[:digit:]	Hyphens and dashes	\p{dash_punctuation}
White Space	Text	GREP	Any Digit	^9	^d	Any lowercase character a to z	[:lower:]	Opening parentheses, braces	\p{open_punctuation}
Em Space	^m	~m	Any Character Not a Digit	^D	^D	Any punctuation and symbols	[:punct:]	Closing parentheses, brackets	\p{close_punctuation}
En Space	^>	~>	Any Letter	^S	[^l\u]	Any whitespace and break	[:space:]	Opening quotes	\p{initial_punctuation}
Flush Space	^f	~f	Any Character	^?	.	Any uppercase character A to Z	[:upper:]	Closing quotes	\p{final_punctuation}
Hair Space	^	~	Any White Space	^w	^s	Any letter, number, and underscore	[:word:]	Underscore, 203F, 2040, 2054	\p{connector_punctuation}
Nonbreaking Space	^S	~S	Any Character Not a White Space	^S	^S	Any hexadecimal 0-9, a-f, and A-F	[:xdigit:]	Other punctuation	\p{other_punctuation}
Nonbreaking Fixed Width Space	^s	~s	Any Vertical Space (hard and soft returns)	^v	^v	Any character of a certain glyph set (such as a, à, á, â, ã, ä, å, Ä, Å, Æ, Æ, Æ)	[[=a=]]	Other	\p{other}
Ideographic Space	^~	~^	Any Horizontal Space (all but hard returns)	^h	^h	Any blank character, ie space or tab	[[blank:]]	Control characters	\p{control}
Thin Space	^<	~<	Any Word Character	^w	^w	Any control character	[[control:]]	Formatting character markers	\p{format}
Figure Space	^/	~/	Any Character Not a Word Character	^W	^W	Any visible character	[[graph:]]	E000-F8FF	\p{private_use}
Punctuation Space	^.	~.	Any Lowercase Letter	^l	^l	Any visible characters and spaces	[[print:]]	Unassigned Unicode range	\p{unassigned}
Third Space	^3	~3	Any Character Not a Lowercase Letter	^L	^L	Any character code > 255	[[unicode:]]		
Quarter Space	^4	~4	Any Uppercase Letter	^u	^u	(applies only to the wide character traits classes)			
Sixth Space	^%	~%	Any Character Not Uppercase	^U	^U				

GREP and InDesign CC 2015 some useful search examples

1. Find ALL CAPS paragraphs

Find: `\u+ (\l\?r?)$`

Change with formatting: paragraph style

Note: When your untagged text has occasional paragraphs typed in all caps, this script finds two or more uppercase letters followed by a period and a hard return. This is enough to find such paragraphs and tag them. Step 2, run `ChangeCaseofSelectedStyle.jsx`, a script by Dave Saunders of PDSAssoc.com. This converts case.

2. Find ALL CAPS words

Find: `\<[A-Z]{2,}\>`

Change: with formatting: OpenType All Caps

Note: The minimal length {2,} is to prevent it picking up all single capitals (remove if you want those as well). Ideally, you search full caps with this regular expression and replace them with OpenType All Caps. Any other way will need handwork or a script. For international text, you should search for `\<\u{2,}\>` (the `\u` uppercase wildcard also finds capital "Ü")

3. Find capital letters like acronyms

Find: `\u+` or Find: `>\<\u+ \>`

Change: with formatting; such as AllSmallCaps

Note: This finds 2 or more uppercase letters. The `\<` and `\>` are word delimiters, so it won't alter anything else than caps words. If the Replace field is empty and it has formatting, it applies the formatting to the found text. Only if there is also no formatting, it replaces with nothing.

4. Swap List of First Names and Last Names

Find: `(+) (+)`

Change: `$2, $1`

Note: This finds something to the left and right of a space; and puts them back in reversed with a comma and space

5. Thin spaces around em dashes

Find: `(?<!~<)(~_)(?!~<)`

Change: `~<$0~<`

This finds em dashes without thin spaces and puts them in.

6. Find and delete soft return line breaks

Find: `\s?n`

Change: with single spacebar space

Note: Soft returns replaced by spacebar space

7. Find quoted words and change to italic

Find: `(~\{)(\u*\l*\s*\w*\d*)(~\})`

Change: `$2` and italic character style formatting

Note: This finds the quotes but leaves them out since it only changes to found group 2; not 1 and 3.

8. Find all email addresses

Find: `[\u\d_%-]+@[\u\d_%-]+`

Change: with formatting

Note: The `_`, `%`, and hyphen were added because those characters are not included in the wildcard codes. The square brackets group the codes together. The `+` indicates that these characters can appear any number of times, or not at all. The `@` sign is outside the group to appear only once. The group is then repeated for the rest of the address.

9. Capitalize first letter of word after bullet

Find: `(.)(\b)(\l)`

Change: `$1$2$3`

Note: In the Replace formatting set text to uppercase

10. Search for words with brackets

Find: `\{\w{4,20}\}`

Change: with formatting

Note: This finds with brackets (could be parentheses)

11. Search for words within brackets

Find: `(?<=[\])\w{4,20}(?=[\])`

Change: with formatting

Note: This finds within but not including brackets

12. Search for phone numbers to regularize

Find: `\d?[-.]?\(\d{3}\)?[-.]?\(\d{3}\)?[-.]?\(\d{4}\)`

Change: `$1.$2.$3`

Or Change: `$1-$2-$3` or Change: `($1) $2-$3`

Note: This finds a sequence of 3, 3, and 4 digits. The `?` symbol indicates that the element can exist zero or one time in the string, so the code `[-.]?` means that a hyphen, dot, or space may or may not exist between the digits. Replace can use periods, parentheses, or hyphens, as you wish.

13. Search for dates to format

Find: `\d\d?\- \d\d?\- (\d\d)? \d\d`

Change: with formatting from the character styles panel

Notes: This sequence finds dates formatted like 09-04-2008 as well as 9-4-08, and allows you to apply a character style.

14. Replace multiple terms with one term

Find: `illustration | graph | map | chart`

Change: figure

Note: The list of alternative words must be separated by the vertical bar, and each alternative is replaced with the replace term. GREP is case sensitive by default; so to replace case-insensitively, add `(?)` before the expression.

15. Find or remove trailing characters after numbers, as in 1st, 2nd, 3rd, 4th

Find: `(?<=\d)+st|(?<=\d)+th|(?<=\d)+nd|(?<=\d)+rd`

Change: leave blank, to strip out; or apply formatting of superscript character style

Note: In this example, we are looking for the two-letter trailing characters to strip off the number. The pipe character is an OR statement. So, the search repeats itself four times, saying that if the characters occur where there is a digit preceding it, then the characters will be removed. If you surround the text with parentheses, you could superscript the found text instead of stripping it out.

16. TitleCase applied except for small words

Find: `:\s\l[-\s]\l(?:\f|he|nd?|o|y|rom|or|ut)?\b`

Change: formatting to All Caps styling

Note: Search the selection. Leave the Change what field blank. This won't work on all uppercase text! This also converts to uppercase any word that follows a colon or a hyphen.

17. Undocumented negation in Wildcards

Not documented in the Wildcards flyout is that the uppercase version of the wildcards negates them.

Note: Thus, `\D` finds everything that is not a digit; `\U` finds everything that is not an uppercase letter, and `\L` finds everything that is not a lowercase letter, etc.

18. Find underlining and replace with italics

Find Formatting: Underline

Change Formatting: Italic character style

This uses Text; not GREP, and is good because character styles control better than direct formatting. Note: Have pre-built character styles for bold, italic, and bold italic. Also, have superscript, smallcaps, and other character styles defining swatch colors pre-built for use in Find/Change as well as nested styles.

19. Interpret GREP into English

RegexBuddy (<http://www.regexbuddy.com/>)

Notes: To see the translation, click the Create tab after entering a regular expression. You will see that the program has listed your expression as a sequence of icons/explanations. Windows-only.

Michael Witherell
mikewitherell@jetsetcom.net

Urszula Witherell
urszulawitherell@jetsetcom.net

Join us for Adobe Certified Expert training in:
InDesign | Illustrator | Photoshop | Acrobat | InCopy

Book us online at: www.jetsetcom.net
and www.trainingsite.com