

GREP and Metacharacters in the Find/Change dialog box of InDesign CS5 & 5.5

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

GREP and InDesign CS5 some useful search examples to explore

1. Find ALL CAPS paragraphs

Find: `\u\u+(\.\\?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 "Ū")

3. Find capital letters like acronyms

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

Change: with single spacebar space

Note: Soft returns replaced by spacebar space

7. Find quoted words and change to italic

Find: `(~){(\u*\|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: `(.)(\t)(\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\\|[-s\\|\\!(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.trainingonsite.com