

# GREP and Metacharacters in the InDesign CS6 Find/Change dialog

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

# GREP and InDesign CS6 some useful search examples

## 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 "ü")

## 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?\\n`

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\\(\\)`

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|h|e|n|d|?|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](mailto:mikewitherell@jetsetcom.net)

Ursula Witherell | [ursulawitherell@jetsetcom.net](mailto:ursulawitherell@jetsetcom.net)

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

Book us online at: [www.jetsetcom.net](http://www.jetsetcom.net)  
and [www.trainingonsite.com](http://www.trainingonsite.com)