

By Jitendra Narayan, BioinformaticsOnline.com

Regular Expression is the heart of Perl Programming; this Cheat List is being prepared in order to help the biologist working on huge Genomes and Protein sequences.

I need you valuable comments on info@bioinformaticsonline.com

Anchors

^	Start of line
\A	Start of string
\$	End of line
\Z	End of string
\b	Word boundary
\B	Not word boundary
\<	Start of word
\>	End of word

Character Classes

\c	Control character
\s	White space
\S	Not white space
\d	Digit
\D	Not digit
\w	Word
\W	Not word
\xhh	Hexadecimal character hh
\Oxxx	Octal character xxx

Control character

[:upper:]	Upper case letters
[:lower:]	Lower case letters
[:alpha:]	All letters
[:alnum:]	Digits and letters
[:digit:]	Digits
[:xdigit:]	Hexadecimal digits
[:punct:]	Punctuation
[:blank:]	Space and tab
[:space:]	Blank characters
[:cntrl:]	Control characters
[:graph:]	Printed characters
[:print:]	Printed characters and spaces
[:word:]	Digits, letters and underscore

By Jitendra Narayan, BioinformaticsOnline.com

--	--

Assertions

<p> ?= ?! ?<= ?!= or ?<! ?> ?() ?() ?# </p>	<p> Lookahead assertion + Negative lookahead + Lookbehind assertion + Negative lookbehind + Once-only Subexpression Condition [if then] Condition [if then else] Comment </p>
--	--

Sample Patterns

<p> ([A-Za-z0-9-]+) (\d{1,2}\.\\d{1,2}\.\\d{4}) (^\s+(?=\.(jpg gif png))\.\2) (^([1-9]{1}\$ ^([1-4]{1}[0-9]{1}\$ ^50\$)) (#?([A-Fa-f0-9]){3}(([A-Fa-f0-9]){3})?) ((?=\.*\d)(?=\.*[a-z])(?=\.*[A-Z]).{8,15}) (\w+@[a-zA-Z_]+?\.\[a-zA-Z]{2,6}) (\<(/?[^\>]+)\>) </p>	<p> Letters, numbers and hyphens Date (e.g. 21/3/2006) jpg, gif or png image Any number from 1 to 50 inclusive Valid hexadecimal colour code 8 to 15 character string with at least one upper case letter, one lower case letter, and one digit (useful for passwords). Email addresses HTML Tags </p>
---	---

Quantifiers

<p> * *? + +? ? ?? {3} {3,} {3,5} {3,5}? </p>	<p> 0 or more + 0 or more, ungreedy + 1 or more + 1 or more, ungreedy + 0 or 1 + 0 or 1, ungreedy + Exactly 3 + 3 or more + 3, 4 or 5 + 3, 4 or 5, ungreedy + </p>
--	---

Ranges

<p> . (a b) (...) </p>	<p> Any character except new line (\n) + a or b + </p>
--	--

By Jitendra Narayan, BioinformaticsOnline.com

(?...) [abc] [^abc] [a-q] [A-Q] [0-7] \n	Group + Passive Group + Range (a or b or c) + Not a or b or c + Letter between a and q + Upper case letter + between A and Q + Digit between 0 and 7 + nth group/subpattern +
--	---

Special Characters

\ \n \r \t \v \f \a [\b] \e \N{name}	Escape Character + New line + Carriage return + Tab + Vertical tab + Form feed + Alarm Backspace Escape Named Character
---	--

Pattern Modifiers

G i m s x e U	Global match Case-insensitive Multiple lines Treat string as single line Allow comments and white space in pattern Evaluate replacement Ungreedy pattern
---------------------------------	---

Special Characters

\ \n \r \t \v \f \a [\b] \e	Escape Character + New line + Carriage return + Tab + Vertical tab + Form feed + Alarm Backspace Escape
---	---

By Jitendra Narayan, BioinformaticsOnline.com

\N{name}	Named Character
----------	-----------------

String Replacement (Back references)

\$n \$2 \$1 \$` \$' \$+ \$& \$_ \$\$	nth non-passive group "xyz" in /^abc(xyz)\$/ "xyz" in /^(?:abc)(xyz)\$/ Before matched string After matched string Last matched string Entire matched string Entire input string Literal "\$"
--	---

Metacharacters (must be escaped)

^ \$ () <	[{ \ >	. * + ?
------------------------	-----------------------	------------------

Sorry, but I do not have time to provide free answers to questions about Perl and Bioperl.
 (I'm busy trying to make a living!) So please don't e-mail your questions to me.

Enjoy the beauty of Biological Programming with Perl.