Using Regular Expressions in Policy Patrol

Policy Patrol includes a powerful string-matching tool to allow you to easily find and filter (spam) keywords. Spam often contains word derivatives and special characters to try to fool spam filters. For instance instead of using the word Viagra in a solicitation the offending email may use word variations such as V.I.A.G.R.A, VIAGRA!!, $VIAGRA$, VlAGRAV and V1AGRA to fool spam filtering software. With Policy Patrol you can now easily create one single regular expression to catch all these variations. Policy Patrol can also use regular expressions to detect credit card numbers and Social Security Number sequences.

Regular expressions have been used in Unix systems and scripting languages like Pearl and Python for many years. They can be simple or very complex. There are many resources including entire books written on using regular expressions (see more information). This document describes the basic Regular Expression syntax and how to use Regular Expressions in Policy Patrol.

Regular Expression Syntax

The following is a list of some of the most common regular expression syntax. You should be careful to not create overly general expressions. Although `.*hot.*` would match many spam phrase like 'hot offer' or 'hot deal on prescription drugs', it would also match 'hotel' and 'photography', which are not necessarily spam words.

Character	Function	Description/Example
.	*Any single character*	The dot (.) Matches any single character. *Examples:* 'me.' would match men, met, etc (but not me as it requires an additional character)

*	*Zero or more of the previous character*	The asterisk (*) matches zero or more instances of the previous character in order. *Examples:* 'mo*' would match zero or more of the character 'o', e.g. mo, moo, and mooo. 'fa.*' would match zero or more of the special character (.), i.e. any single character. So it would match words such as fa, fan, and fantastic.
+	*One or more of the previous character*	The plus sign (+) matches 1 or more of the previous character *Examples:* *'mo+'* would match moo or mooo (but *not* mo as it requires 1)
?	*Zero or one of the previous character*	The question mark (?) matches 1 or 0 of the previous character *Examples:* *'mi?'* would match mi or mii (but *not* miii as it matches only 0 or 1).
[]	*Any character from the set*	The square brackets ([]) match any character from a predefined group. *Examples:* '[aeiou]' would match any vowel.
[^]	*Any character not from the set*	The circumflex characters is a not operator, inside square brackets ([^]) it would match any character not from a predefined group. *Examples:* '[^aeiou]' would match any character not a vowel (any consonant).
|	*Or*	The separator | can for instance be used with the group characters (and). Examples: '(a|@)' would match a and @. '(box|filter|tv)' would match box, filter and tv.

\s	*White space*	Backslash s matches any white spaces including a tab. Examples: 'reduce\sdebt' would match reduce debt, but not reduce$debt, or reducedebt.
{ *num* }	The preceding element *num* times	A number within curly braces matches the preceding element that number of times. *Examples:* '[aeiouy]{3}' would match any word with three vowels in a row.
{*min*, *max*}	The preceding element between *min* and *max* times	Two numbers (the second must be >= the first), separated by a comma, within curly braces matches the preceding element between *min* and *max* times. *Examples:* '[aeiouy]{2,5}' would match any word with between two and five vowels in a row. '[aeiouy]{2,}' would match any word with more than two vowels in a row. '[aeiouy]{,3}' would match any word with fewer than 3 vowels in a row.
\b	The *start* of a word	To match whole words only, use the character sequence "\b". *Examples:* '\bhot' would match the word hot and hotel, but not photo. '\bhot\b' would match only hot. *Note:* Remember that the Policy Patrol filter options 'Whole word(s) are matched' and 'Whole or part of word(s) are matched' are not operative with regular expressions. Use the \b sequence at the beginning and end of your regular expression if you do not want to match word subsets.

Regular Expression examples

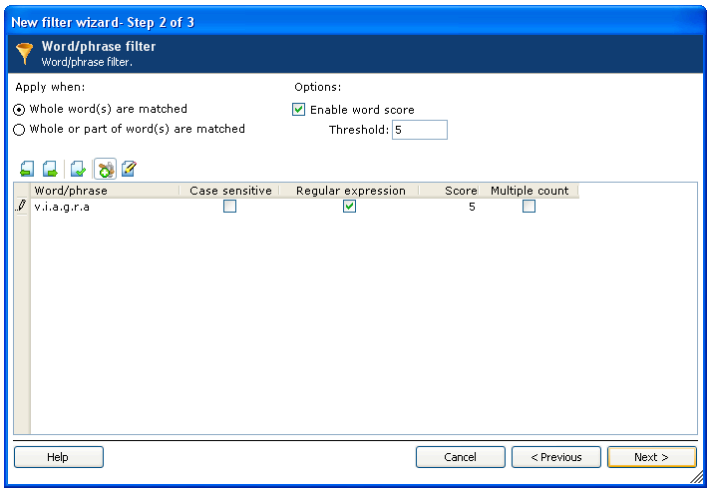
A few regular expression examples are listed below.

Regular Expression	Matches
v.i.a.g.r.a	This would match v*i*a*g*r*a and v/i/a/g/r/a, but not viagra or vi@gra.
buy.*now	This would match buy now, buy.now, buy-now, buy//now and buynow.
d.?e.?b.?t	This would match debt, d*e*b*t and d-e-b-t but not d//e//b//t.
\$\$+	This would match $$$$$, $$ and $$!! but not $.
'…V..[A|@]..G..R..[A|@]…'	This would match a word with any first three letters, followed by a V, followed by any two characters, then an A or an @, etc, such as: V.I.A.G.R.A, VIAGRA!!, $VIAGRA$, ,V1AGRA, VI@GR@.
'...V..A..G..R…A...'	This would match V.I.A.G.R.A, VIAGRA!!, $VIAGRA$, VlAGRAV and V1AGRA, but not VI@GR@.
'[^aeiouy]{6,}'	This would match words with more than 6 consonants in a sequence.
'[^ \t\n\r\f]{35,}'	This would match words with more than 35 characters.
cable.?.?.?.?.?(box|filter|tv).?	This would match: cabletv, cable*tv, cablebox, cable-box, freecable tv and cable*****box*.

Using Regular Expressions in Policy Patrol

Regular Expressions can be entered in Policy Patrol Word/phrase filters or in **Email Threat Protection** > **Block > Word Block List** or in **Email Threat Protection** > **Allow > Word Allow List**. To use regular expressions, you must enter the regular expression in the Word/phrase column and tick the box regular expression.

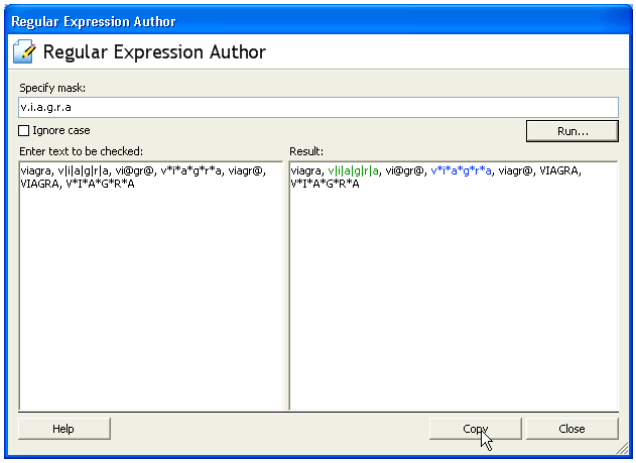
Note: The options **Whole word(s) are matched** and **Whole or part of word(s) are matched** do not apply to regular expressions since this can be indicated in the regular expression itself (see '\b' in the Regular Expression Syntax). The case sensitivity and multiple count options are valid for regular expressions.



Regular Expression Author

Policy Patrol includes a Regular Expression Author to help you create and test your regular expressions. Follow the next steps to use the Regular Expression Author:

1. In your Word/phrase filter, click on the **Regular Expression Author** icon in the toolbar.



2. In **Specify mask**, enter your regular expression, for instance `v.i.a.g.r.a`. If you wish to ignore case, select the option **Ignore case**.

3. In the left dialog, enter the sample text to be checked for the regular expression.

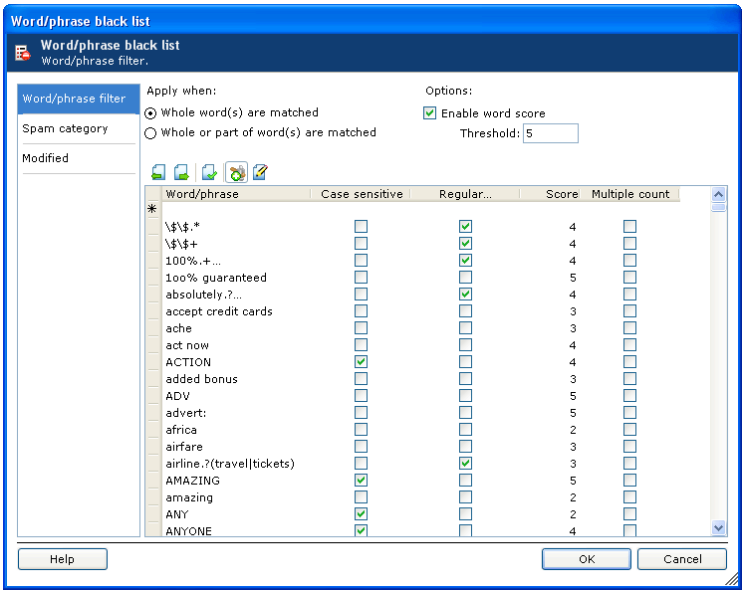
4. Click on **Run**. The words that match the regular expression will be colored green and blue alternately. For instance, in the example above, you can see that the regular expression v.i.a.g.r.a matches v*i*a*g*r*a, but not viagra or vi@gra.

5. If the result is not as you had intended, alter the regular expression and press **Run** again. If your regular expression produced the intended results, press **Copy** and **Close**. Now paste the regular expression into the word/phrase filter and tick the box **Regular expression**.

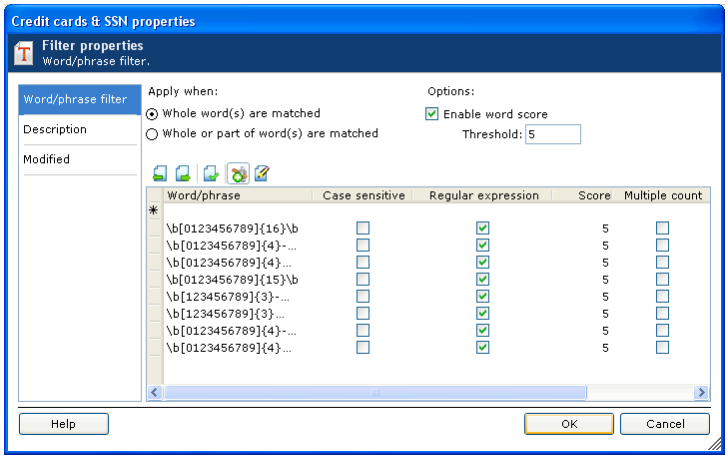
Note: Be cautious when using the * sign in word entries. If the word is not marked as a regular expression, the * is seen as a wildcard for any character. This means that if you enter the word v*i*a*g*r*a this will not only find v/i/a/g/r/a and v-i-a-g-r-a, but also the phrase: Victor is a great person. If you enter the word v*i*a*g*r*a and check the regular expression tick box, this means that the entry will trigger on all words since the * sign means 0 or more of the previous character.

Sample Regular Expression filters

Policy Patrol includes a sample word/phrase black list that makes use of regular expressions to effectively catch spam.



Policy Patrol also ships with a sample word/phrase filter that detects credit card numbers and Social Security Number sequences.



Need assistance?

Do you need assistance in creating your regular expression? Please send a description of the words you would like to detect to support@redearthsoftware.com and we will email you back the regular expression that you can use.

More information

⇒ For more information on regular expressions, we recommend the following books on the subject: 'Mastering Regular Expressions', http://www.oreilly.com/catalog/regex/ and 'Sams Teach Yourself Regular Expressions in 10 Minutes', http://www.forta.com/books/0672325667/.

⇒ For more information on how to configure Policy Patrol, please consult the program help or download the product manual from: http://www.policypatrol.com/documentation.

⇒ For more information on how to configure word/phrase filtering in Policy Patrol, please download the document 'Word/phrase filtering with Policy Patrol' from: http://www.policypatrol.com/assets/pp10-wordfiltering.pdf.

⇒ If you still have questions after reading this document, please consult our online knowledge base at http://support.redearthsoftware.com, or send an email to support@redearthsoftware.com.

Contacting Red Earth Software

Please contact us at one of the following offices:

Red Earth Software, Inc.
4845 Pearl East Circle, Ste 101
Boulder, CO 80301
United States
Toll-free: 1-800-921-8215
Phone: (720) 377-3728
Fax: (720) 554-7950
Sales: sales@redearthsoftware.com

Red Earth Software (UK) Ltd
20 Market Place
Kingston-upon-Thames
Surrey KT1 1JP
United Kingdom
Tel: +44-(0)20-8328 9830
Fax: +44-(0)20-8711 5771
Sales: sales@redearthsoftware.co.uk

Support: support@redearthsoftware.com Support: support@redearthsoftware.co.uk

Policy Patrol® is a registered trademark of Red Earth Software®. Copyright © 2001- 2014 by Red Earth Software.