

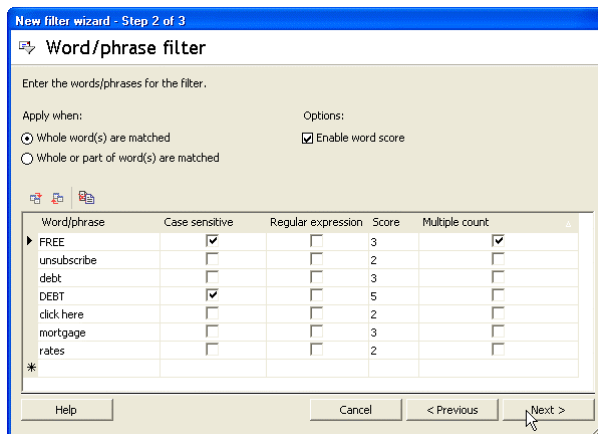
## Word/phrase filtering with Policy Patrol

Policy Patrol can search emails for words and phrases by making use of word/phrase filters. In order to increase accuracy and avoid false positives, Policy Patrol allows you to specify word score, case sensitivity and multiple count. In addition Policy Patrol offers word pattern matching through the use of regular expressions.

### Creating a Word/Phrase Filter

Follow the next steps to create a Word/Phrase filter:

1. Go to **Filters** > **<folder>** and click **New....**
2. When asked which type of filter you wish to create, select **Word/phrase Filter**. Click **Next**.
3. Enter the word(s) or phrases to be included in the filter. For each entry you can apply a word score, and select whether it should be case sensitive and whether it should be counted multiple times.

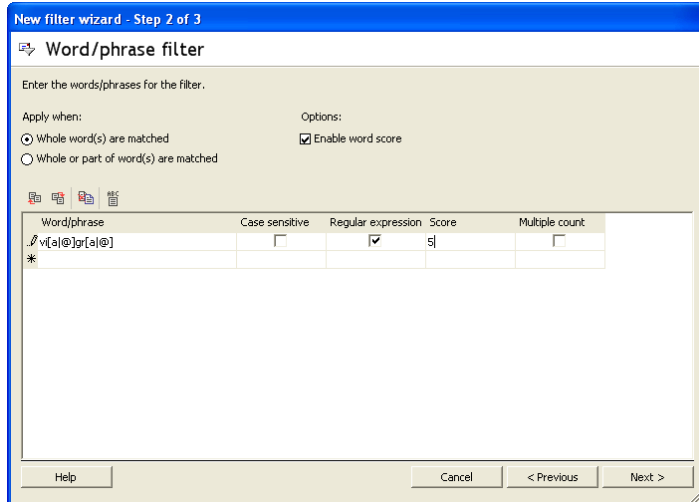


#### Case sensitivity

If you check the **Case sensitive** option, this means that Policy Patrol will only check for the word in the same case. This can be useful for certain spam or chain letters for instance, that tend to use a lot of capitals. For instance if a mail includes CLICK HERE in capitals there will be a good chance that the mail is spam. However, click here in lower case might be more innocent. By using the case sensitive option in combination with the word score option you could add both variations, applying a higher score to the upper case version. Remember though that if you enter a case sensitive and non-case sensitive version of the same word, and the word in the email matches the case sensitive version, the word will be counted twice since it will match both the case sensitive and non-case sensitive entry in the filter.

## Regular expression

If the entry is a regular expression tick the box **Regular expression**. Regular expressions allow you to match a word pattern instead of an exact word. This means that by making use of regular expressions you can stop spammers trying to circumvent content filters by adding characters within words, such as v\*i\*a\*g\*r\*a or c-l-i-c-k h-e-r-e. Furthermore you can detect word variations such as r@tes and l0ans. Policy Patrol includes an extensive spam words filter that makes use of many regular expressions to detect variations of spam words.



Below are a couple of expressions that can be used:

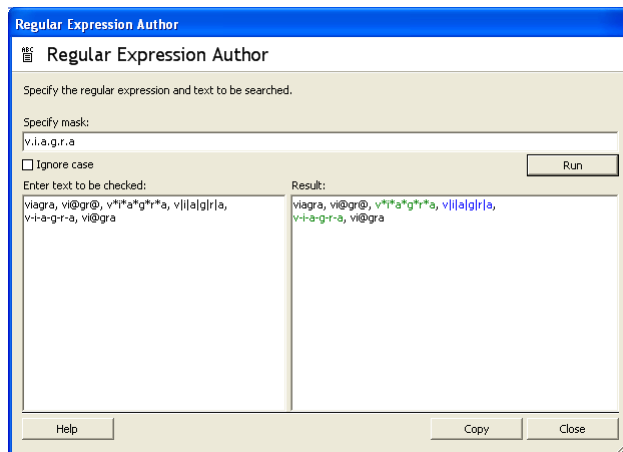
Regular expression	Meaning
\b	Word boundary
.	Any character
*	Previous character 0 or more times
+	Previous character 1 or more times
[a,b]	Character a or b

For instance, if you enter `vi[a|@]gr[a|@]` this will find the word `viagra`, `vi@gra`, `viagr@` and `vi@gr@`. If you enter `v.i.a.g.r.a`, this will find the words `v*i*a*g*r*a`, `v|i|a|g|r|a` and `v-i-a-g-r-a`. Note that 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. The case sensitivity, word score and multiple count options do apply.

**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: `V`ictor is `a` `g`reat 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.

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. 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**.

### **Word score**

If you wish to use word scores in the filter, check the option **Enable word score** and enter the score for each entry. When configuring the rule, you will specify the word score threshold that must be met in order to trigger the rule. If the total score for the words found in the message equals or exceeds the word score threshold, the rule will trigger. In other words, if you enter two words in the filter with both a score of 5, and the rule is configured with a word score threshold of 10, the rule will trigger when at least both entries are found in the email or, if multiple count is ticked, two instances are found of one of the entries. If you do not wish to use word scores in the filter, uncheck **Enable word score**.

### **Negative word score**

You can also apply a negative word score. This can be useful to eliminate some words that can be used innocently. For instance you might assign the word 'breast' a word score of 5, and assign the words 'baby' or 'chicken' a minus 5 score. You can also add a negative score to words that indicate legitimate emails such as your company name and your product or service name. By setting different word scores and applying negative scores for certain words, it is possible to closely identify the content of emails and in doing so greatly decrease the occurrence of false positives (i.e. wrongly triggered rules).

### ***Multiple count***

If you wish every instance of the word to be counted, check the box **Multiple count**. For example, if this box is enabled and you receive an email message that contains the word 'debt' three times, and you applied word score of 5 to this word, the total word score would be 15. If you did not check this box, the word will only be counted once and the total score would be 5.

### ***Import/Export***

You can import lists from .txt files by clicking on **Import**, browsing to the appropriate file and clicking **Open**. The format should be as follows: word/phrase;case sensitive; score; count multiple; regular expression. The word/phrase and score values must be entered. For the other options, either TRUE or FALSE must be entered. For instance, if you wish to add the case sensitive word CLICK HERE with a word score of 5 and multiple count, you must enter it in the text file as follows: CLICK HERE;TRUE;5;TRUE;FALSE. For every word or phrase you need to start a new line. If you import words or phrases from more than one file, the additional words or phrases will be added to the list. If you have two lists with some common words, Policy Patrol will not add the common words twice, but will only add the additional ones. To export the words in the filter, click **Export**, enter a file name and select **OK**.

### ***Whole or part of words***

Select whether to apply the filter when **Whole word(s) are matched** or when **Whole or part of word(s) are matched**. The first option allows you to specify more precisely which words must trigger a rule. For instance, if you select that **Whole or part of word(s) are matched** and you enter the word 'sex' in the filter, this will also include the words 'Sussex' and 'sextant'. If you select **Whole word(s) are matched**, the rule will trigger on the word 'sex' but not on 'Middlesex'.

When you are ready adding words, click **Next**.

4. Enter a name for the filter and any additional comments. When you are done, click **Finish** to create the filter.

## Configuring word/phrase filtering in the rule

---

Follow the next steps to configure a rule that checks for words or phrases:

1. Go to **Rules** > **<folder>** and click **New**.
2. Select the users for the rule and click **Next**.
3. Select which messages you want Policy Patrol to check and click **Next**.
4. Select **Trigger rule if following conditions are met**. For word/phrase filtering you can select one or more of the following conditions:
  - **Subject contains word/phrase** (searches the subject of the email)
  - **Body contains word/phrase** (searches the body of the email)
  - **Attachment contains word/phrase** (searches the attachment of the email)
5. After selecting one of the above options, click on the word/phrase link in the description and select the filter(s) you want to check.

If you selected to search the body and you wish to search for HTML tags, check the option **Check HTML tags**. This can be useful if you want to check for scripts by searching for the `<SCRIPT>` tag. However, if you wish to check normal text, you must not select this option since this will produce unwanted results.

6. When you select a word/phrase option, the **Word score threshold** box will become active (situated below the description pane). Here you must enter the total word score for the email message that should trigger the rule. If the total score of words/phrases found in an email is equal to or above the word score threshold, the rule will trigger. For instance, if the email message contains the phrase 'CLICK HERE', which is attributed a word score of 5 and the rule has a word score threshold of 5, the rule will trigger.

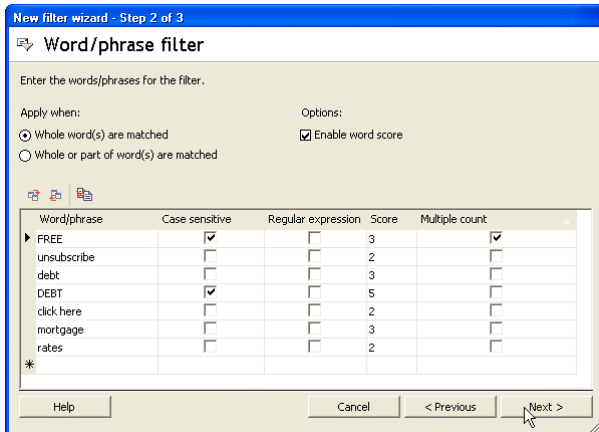
If you specify more than one word/phrase filter to content check, Policy Patrol will add the scores of all words/phrases and trigger the rule once the word score threshold is reached. Similarly, if you select to check word/phrase filters for subject, body and/or attachment, Policy Patrol will add all the scores and trigger once the total score reaches the word score threshold.

If you do not use the word score option in your word/phrase filters, you can leave the word score threshold set to '0'. The rule will then trigger when any word/phrase from the filter is found. If you configure a rule that checks multiple word/phrase filters, some with and some without word score enabled, Policy Patrol will trigger the rule if words/phrases from the word score filter reach the word score threshold, or if any word/phrase from the non-word score filter is found in the email.

Note that if word score is enabled in the selected filter(s) (which it is by default) you must enter a word score threshold above 0. If you leave the threshold set to 0, the rule will never trigger since a threshold of 0 is considered invalid.

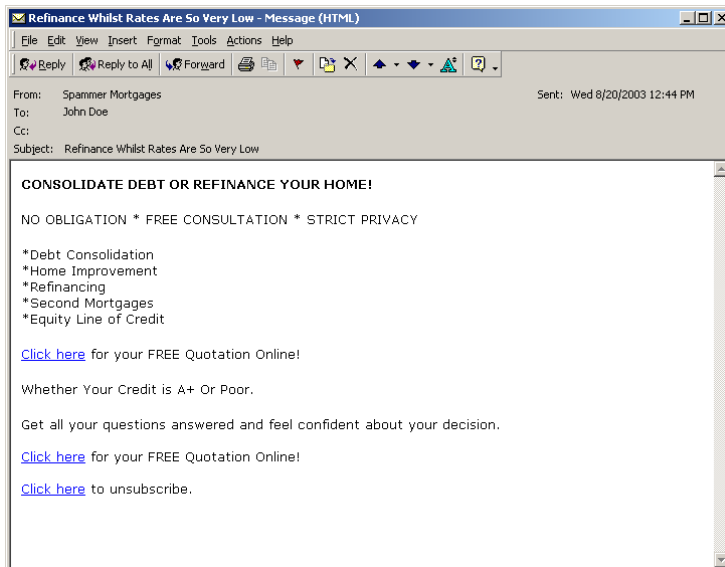
## Example

In order to explain the word score functionality, an example is discussed here. In the example you have configured a Word/Phrase filter 'Spam words' as shown in the screen below. You have selected that whole word(s) should be matched and you have enabled word score.

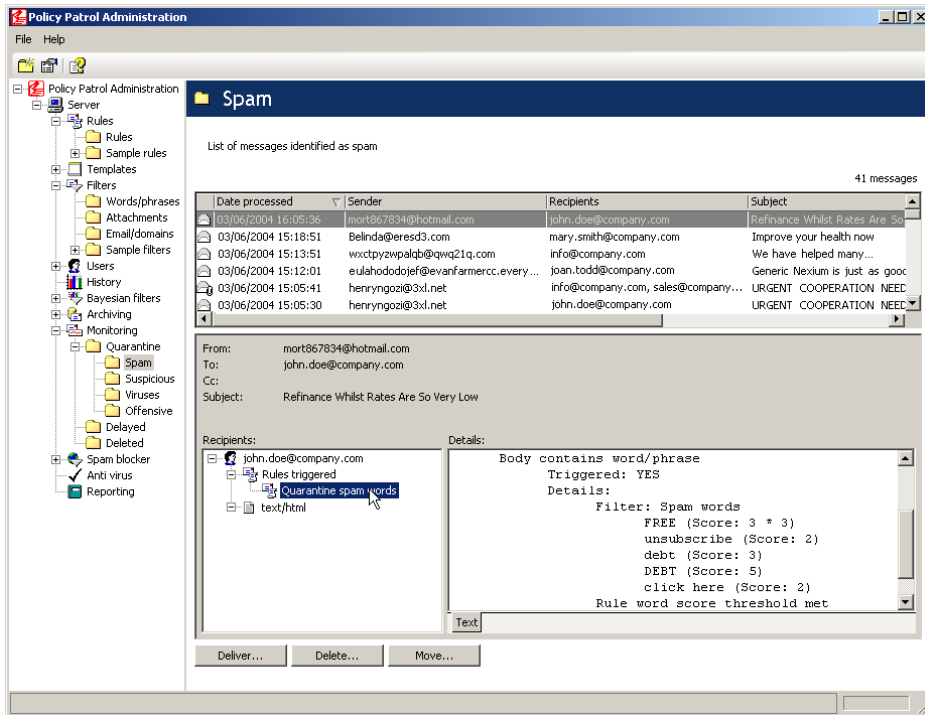


You have also created a rule that checks the email body for words from the 'Spam words' Word/Phrase filter and have applied a word score threshold of 5. This means that the rule will trigger when the total score of words found in the body equals 5 or more.

A spam message is sent to your organization as shown in the screen below.



Policy Patrol quarantines the message. You go to **Monitoring > Quarantine > Spam** and select the quarantined message. When you click on **Rules triggered** and select the rule **Quarantine spam words**, you see which conditions triggered the rule. The words that Policy Patrol found are listed with a total score of 21. Since you configured a word score threshold of 5, the 'Quarantined spam words' rule triggered.



We now take a closer look at how Policy Patrol determined the word score, by examining the email message (the message is displayed below with the words from the filter in bold).

- ⇒ Although the non case sensitive word `rates` is found in the subject, the rule was not configured to search the subject for words from the 'Spam words' filter. (score=0)
- ⇒ Policy Patrol finds three instances of the case sensitive word `FREE`. Since multiple count is selected for this word, the word score is counted 3 times. (score=9)
- ⇒ One instance of the non case sensitive word `unsubscribe` is found. (score=2)
- ⇒ Policy Patrol finds two instances of the non case sensitive word `debt`. However since multiple count is not enabled for this word, it is only counted once. (score=3)
- ⇒ One instance of the case sensitive word `DEBT` is found (Note that if the same word matches multiple word/phrase entries in the filter, each entry will be counted). (score=5)
- ⇒ Three instances of the non case sensitive word `click here` are found in the email body. However since multiple count is not enabled, the word is only counted once. (score=2)
- ⇒ Policy Patrol does not find the word `mortgage` in the email, since you checked the option **Whole word(s) are matched** in the Spam words filter, and the email message contains the word `Mortgages`. If you had selected **Whole or part of word(s) are matched**, Policy Patrol would have counted this word as well. This

option must be used with caution though, since it can result in some false positives. For instance, if you enter the word `sex` in a Word/Phrase filter and you enable the option **Whole or part of word(s) are matched**, Policy Patrol will also find the word `sex` in words such as Sussex, Middlesex and sextant.

Subject:	Refinance Whilst Rates Are So Very Low
Body:	CONSOLIDATE <b>DEBT</b> OR REFINANCE YOUR HOME!  NO OBLIGATION * <b>FREE</b> CONSULTATION * STRICT PRIVACY  * <b>Debt</b> Consolidation *Home Improvement *Refinancing *Second Mortgages *Equity Line of Credit  <b>Click here</b> for your <b>FREE</b> Quotation Online!  Whether Your Credit is A+ Or Poor  Get all your questions answered and feel confident about your decision.  <b>Click here</b> for your <b>FREE</b> Quotation Online!  <b>Click here</b> to <b>unsubscribe</b> .

## Troubleshooting

---

Q: My rule that searches for words/phrases never triggers

A: Check whether you have enabled word score in the selected Word/Phrase filter(s), and have left the word score threshold in the rule at 0. In this case the rule will never trigger since a threshold of 0 is considered invalid.

Q: My rule that searches for words/phrases is triggering erratically

A: Check whether the option **Check HTML tags** is selected in Rule Properties > Conditions tab > word/phrase filter link. If this option is selected Policy Patrol will search HTML tags as well as text and might produce unwanted results if used for checking normal text. For instance, spammers frequently use comment tags within the text (which are not displayed on screen) to circumvent content filters. By default, Policy Patrol will ignore these tags, however if **Check HTML tags** is selected it will include these tags in the search.

## More information

---

⇒ For more information on how to use regular expressions, please download the document 'Using regular expressions in Policy Patrol' from:  
<http://www.policypatrol.com/docs/PP3-RegularExpressions.pdf>.

- ⇒ For more information on how to configure Policy Patrol, please consult the Quick Start guide, program help or download the product manual from: <http://www.policypatrol.com/docs/policypatrol3manual.pdf>.
- ⇒ For frequently asked questions, consult our knowledge base at: <http://www.policypatrol.com/kb.asp>.
- ⇒ If you would like some more assistance on configuring your word/phrase filters and rules, please run the support wizard from Help > Support Wizard or send an email to [support@reearthsoftware.com](mailto:support@reearthsoftware.com).

## Contacting Red Earth Software

---

Red Earth Software LLC  
200 Marcy Street  
Portsmouth, NH 03801  
United States  
Phone: (603) 436-1319  
Fax: (603) 457-8455  
Sales: [sales@reearthsoftware.com](mailto:sales@reearthsoftware.com)  
Support: [support@reearthsoftware.com](mailto:support@reearthsoftware.com)

Red Earth Software (UK) Ltd  
20 Market Place  
Kingston-upon-Thames  
Surrey KT1 1JP  
United Kingdom  
Tel: +44-(0)20-8605 9074  
Fax: +44-(0)20-8605 9075  
Sales: [sales@reearthsoftware.co.uk](mailto:sales@reearthsoftware.co.uk)  
Support: [support@reearthsoftware.co.uk](mailto:support@reearthsoftware.co.uk)

Policy Patrol® is a registered trademark of Red Earth Software®. Copyright © 2001- 2004 by Red Earth Software.