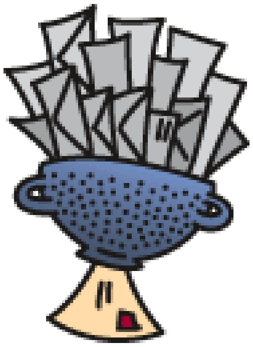
SpamSieve 1.0 Manual

Michael Tsai

September 9, 2002



1 Introduction

1.1 What Is SpamSieve?

SpamSieve is an application that integrates with e-mail clients to filter out unsolicited mass mailings, commonly known as "spam." Previously, most people just ignored spam messages or created simple rules in their e-mail clients to filter it out. In recent years and months, the spam problem has gotten worse. Today's spam is harder to detect, and there is more of it. People have turned to a wide variety of software solutions to help regain control over their inboxes. This software differs primarily along two dimensions: how it identifies spam messages; and how it reduces their burden on the user.

1.2 Identifying Spam

Simple rules that you create in your e-mail client generally look for common spam words. If a message's subject contains "fix your credit" or its body contains "free porn" it's probably spam. You can look at the spam you receive and recognize some patterns. Unfortunately, spam is constantly evolving, so you will have to keep working on your rules. Eventually you get to the point where you don't know what else to add, afraid that adding more words or phrases to your "blacklist" would start filtering out legitimate mail. You can somewhat get around this by creating a "whitelist" that accepts every message from the people you regularly correspond with. However, this doesn't help for messages received from new people, and it's not uncommon for spam messages to contain a forged return address of someone at your own company.

Commercial anti-spam software often combines rules with the notion of a *score*. Rules can look for patterns that make a message more or less likely to be spam. Each rule either increases or decreases the message's spam score. If the score is above a certain threshold, the message is considered to be spam. The flexible nature of the score means that this approach is often an improvement over simple rules. However, the user typically has little or no control over how the spam score is calculated. If a spam message gets through, the user has no recourse but to hope that the next update corrects the problem. Another problem is that spammers can buy this software, too. They can tailor their spam to get through the rules.

SpamSieve uses a statistical technique known as *Bayesian analysis*[1]. It combines the good properties of the above two approaches and adds some of its own. First, you *train* SpamSieve with examples of your good mail and your spam. When you receive a new message, SpamSieve looks at how often its words occur in spam messages vs. good messages. Lots of spammy words means that the message is probably spam. However, the presence of words that are common in your normal e-mail but rare in spam messages can tip the scale the other way. This "fuzzy" approach allows SpamSieve to catch nearly every spam message yet produce very few false positives[2].

Because you train SpamSieve with your own mail, you have full control. If SpamSieve makes a mistake, you can train it with the message in question so that in the future it will do better. Further, since spammers don't have access to the messages you trained SpamSieve with, they have no way of knowing how to change their messages to get through. Whereas other spam filters become less effective as spammers figure out their rules, *SpamSieve becomes more effective over time* because it has a larger corpus of your messages to work from.

1.3 Filtering It Out

Anti-spam software that runs on mail servers filters out spam before you ever see it. This means that unless the filter is perfect, either some spam messages will get through or a few legitimate messages will not. In the first case, you may want additional, client-side, anti-spam software. The second case troubles many people so much that they prefer that the server do no filtering at all.

Other client-side anti-spam software connects to your mail server to delete spam messages before your e-mail client can download them. This works similarly to above, except that to catch all the spam messages you have to run the program right before your regular e-mail program checks for mail. This is difficult to time properly if you check your mail often, and even so you may download some messages that weren't filtered. The anti-spam software may let you see the messages that it filtered out, so that you can verify that there were no false positives. However, you have to do this using its interface, not your e-mail program's (which

[1] For a more in-depth treatment of Bayesian analysis applied spam, see the article by Paul Graham at `http://www.paulgraham.com/spam.html` and the papers it references.

[2] A *false positive* is a good message mistakenly identified as spam. Most users consider false positives to be much worse than *false negatives* (spam messages that the user has to see).

is typically nicer). And if there was a false positive you then have to transfer it into your e-mail program so that you can file and reply to it.

E-mail clients like Entourage have their own spam filters built-in. This is convenient and makes it easy to manually scan for false positives, but there is typically little you can do when the filter makes mistakes.

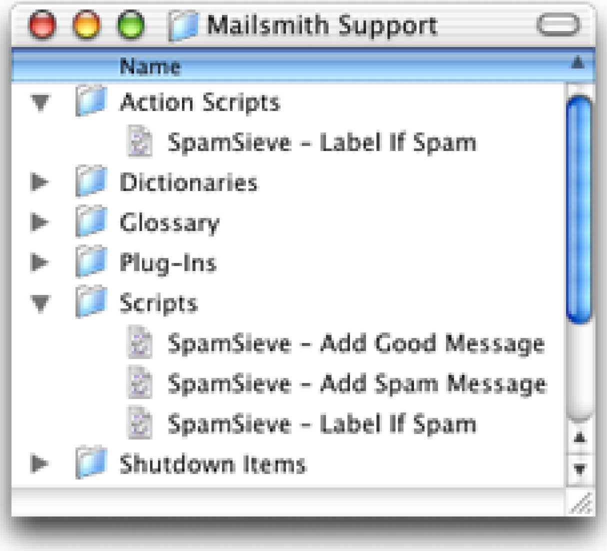
Clearly, the best solution is user-configurable anti-spam software that works directly with your e-mail client. Apple has added exactly this to its Mail program in Mac OS X 10.2. By most accounts this works great. . .unless you prefer a client other than Mail.app. This is where SpamSieve comes in. It brings powerful spam filtering to other popular e-mail clients such as Mailsmith and Entourage.

2 Requirements and Installation

SpamSieve has been developed and tested on Mac OS X 10.1.5 and 10.2. I do not have the resources to test it on older systems, although I suspect it will work fine on Mac OS X 10.1 or later.

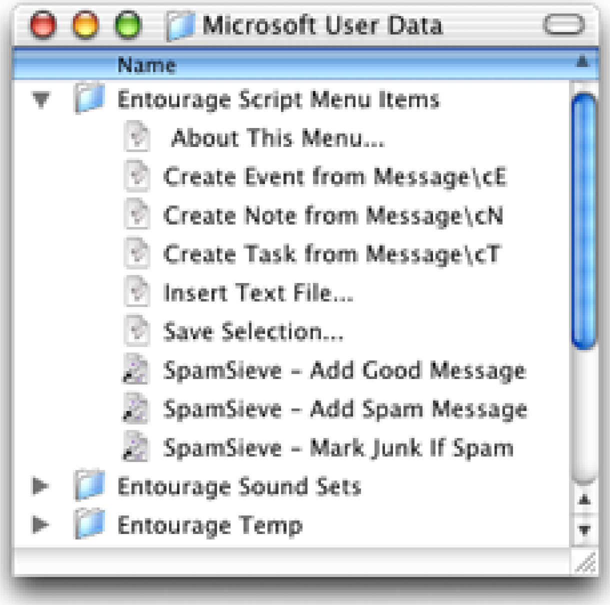
SpamSieve is designed to work with Mailsmith 1.5[3] from Bare Bones Software and Entourage[4] from Microsoft.

To install, copy the SpamSieve application to your hard disk, e.g. into `/Applications`. The names of the other folders tell you where to copy their contents. The AppleScripts in these folders allow you to interact with SpamSieve from within your e-mail client.


Mailsmith Support
Name
Action Scripts
SpamSieve - Label If Spam
Dictionaries
Glossary
Plug-Ins
Scripts
SpamSieve - Add Good Message
SpamSieve - Add Spam Message
SpamSieve - Label If Spam
Shutdown Items

[3] http://www.barebones.com/products/mailsmith.html

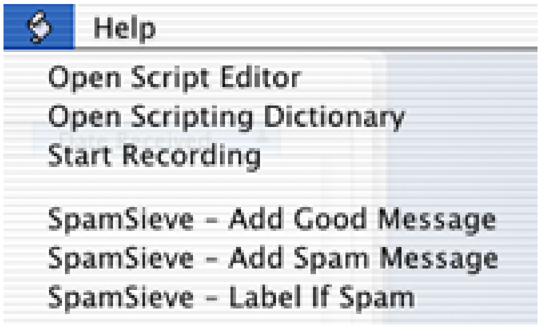
[4] http://www.microsoft.com/mac/entouragex/default.asp?navindex=s4


Microsoft User Data
Name
Entourage Script Menu Items
About This Menu...
Create Event from Message\cE
Create Note from Message\cN
Create Task from Message\cT
Insert Text File...
Save Selection...
SpamSieve - Add Good Message
SpamSieve - Add Spam Message
SpamSieve - Mark Junk If Spam
Entourage Sound Sets
Entourage Temp

There's no need to copy this manual to your hard disk. A copy of it is built into SpamSieve, and you can access it by choosing **SpamSieve Help** from the **Help** menu.

3 Training SpamSieve to Recognize Your Spam

Before you can use SpamSieve, you must give it some examples of messages you consider to be spam, and ones which you do not. Make sure that you have installed the appropriate integration AppleScripts for your e-mail client. Select some spam messages and then use your e-mail client's Scripts menu to run the `SpamSieve - Add Spam Message` script. Then select some good messages and run the `SpamSieve - Add Good Message` script. You can run these scripts at any time, e.g. whenever SpamSieve makes a mistake. The more messages you train SpamSieve with, the better its accuracy it will be. For best results, you should train it with *at least* 600 messages. It is important to train SpamSieve with both spam messsages and good messages. If you can, train it with the same ratio of these two types of messages as you normally receive.


Help
Open Script Editor
Open Scripting Dictionary
Start Recording
SpamSieve - Add Good Message
SpamSieve - Add Spam Message
SpamSieve - Label If Spam

4 Filtering Messages With SpamSieve

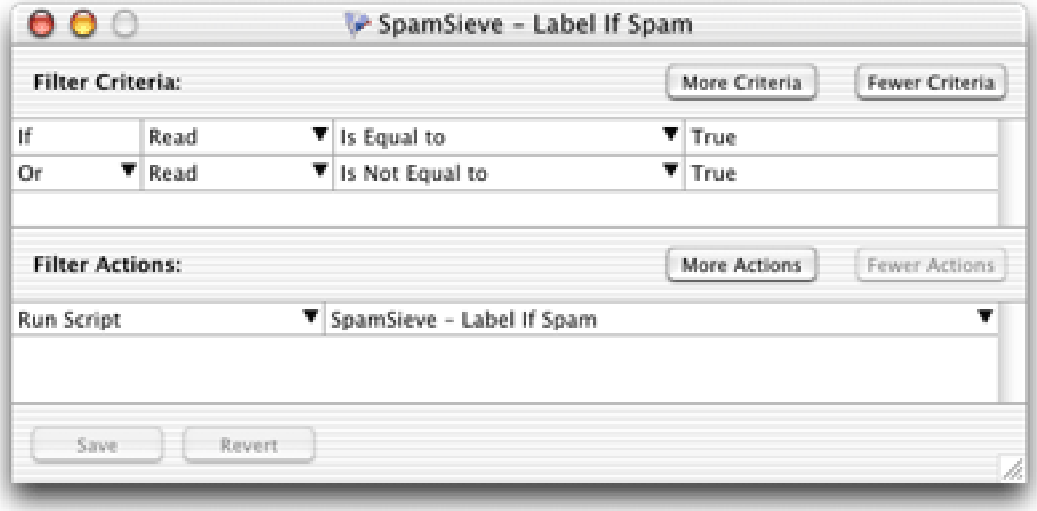
4.1 How SpamSieve Works With E-Mail Clients

Once you've trained SpamSieve, you can begin using it to identify spam messages. The general procedure is that you configure your e-mail client to run an AppleScript each time you receive a message. The AppleScript asks SpamSieve if the message is spam. If it is, it marks the message. You can then set up your own rules in your e-mail client to process the marked messages. You might move them directly to the trash, or into a "Possible Spam" mailbox if you want to manually check for false positives.

Both Mailsmith and Entourage let you control the order in which their rules process mail. How you order the SpamSieve rule is up to you. If you get a lot of spam that happens to match the rules you use to organize your mail, you might want to run the SpamSieve rule first. If you'd rather deal with spam manually than have any false positives, then you might want to run the SpamSieve rule last, after all your other rules have been given a chance to match.

4.2 Setting up a Mailsmith Filter

Create a filter that looks like this:



SpamSieve – Label If Spam

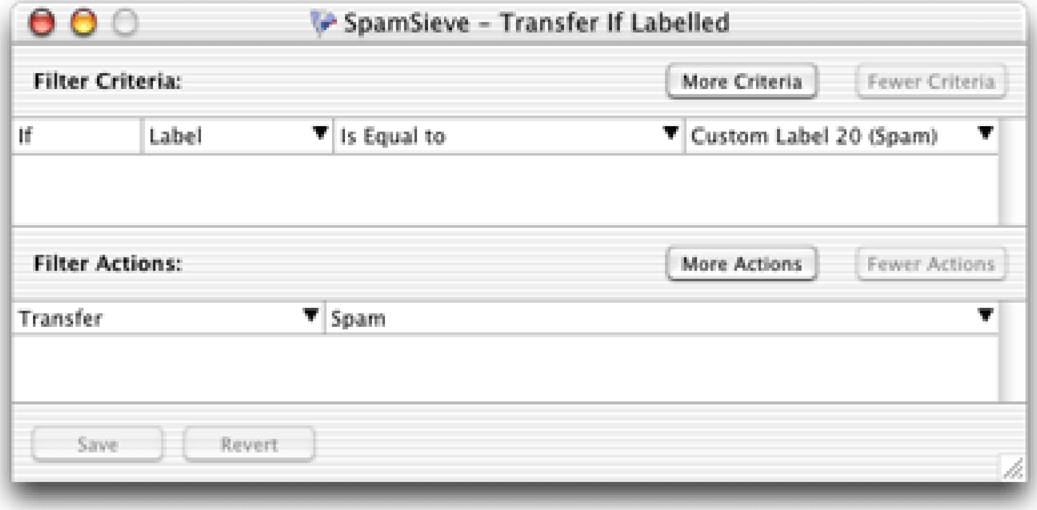
Filter Criteria:

If	Read	Is Equal to	True
Or	Read	Is Not Equal to	True

Filter Actions:

Run Script	SpamSieve – Label If Spam

The two criteria ensure that the filter's action will be applied to every message. The action of the filter sets the label of spam messages to Custom Label 20. You can then use a filter such as the following to transfer spam messages to a particular mailbox:



SpamSieve - Transfer If Labelled

Filter Criteria:

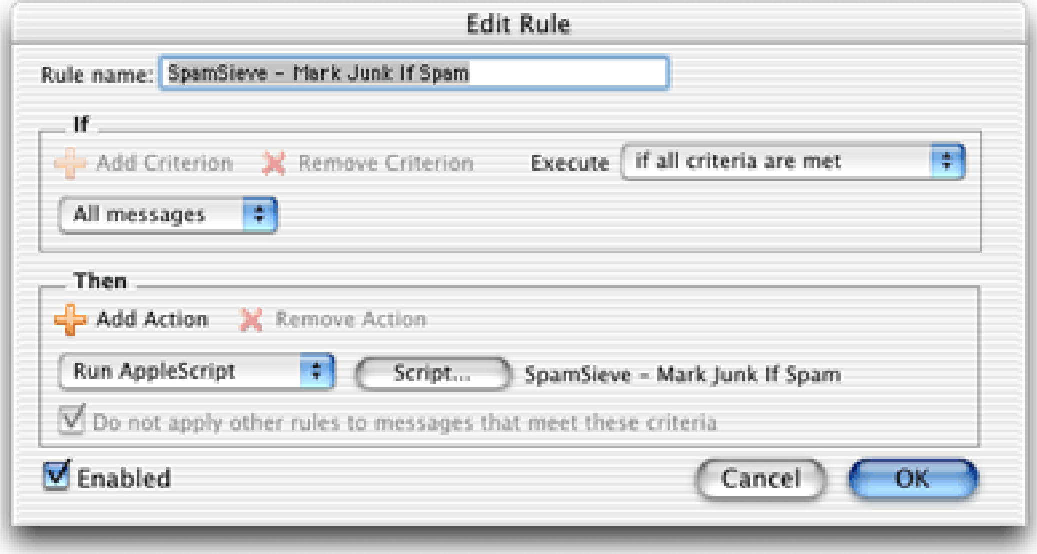
If | Label | Is Equal to | Custom Label 20 (Spam)

Filter Actions:

Transfer | Spam

4.3 Setting up an Entourage Rule

Create a rule that looks like this:



Edit Rule

Rule name: SpamSieve - Mark Junk If Spam

If

Execute: if all criteria are met

All messages

Then

Run AppleScript | Script... | SpamSieve - Mark Junk If Spam

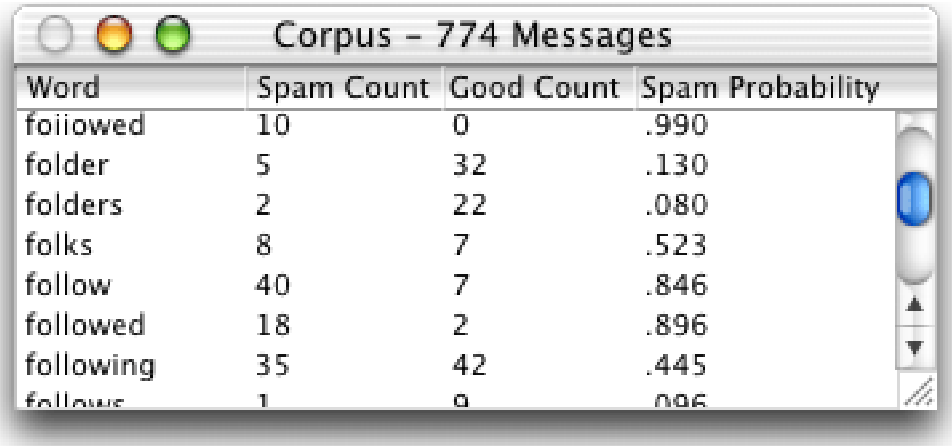
☑ Do not apply other rules to messages that meet these criteria

☑ Enabled

The rule applies to all messages and sets the category of spam messages to Junk. One complication is that after Entourage runs an AppleScript it cannot apply any more rules to the message. However, since the messages will be marked, it is to identify them visually, or with a manually applied rule.

5 Examining the Corpus

The **Corpus** command in SpamSieve's **Window** menu lets you examine the words that SpamSieve has found in your e-mails. It shows how many times each word occurred in spam and non-spam messages, and the probability SpamSieve uses to determine whether a message containing that word is spam. You can click on the name of a column to sort by that column.



Corpus – 774 Messages

Word	Spam Count	Good Count	Spam Probability
foiiowed	10	0	.990
folder	5	32	.130
folders	2	22	.080
folks	8	7	.523
follow	40	7	.846
followed	18	2	.896
following	35	42	.445
follows	1	9	.096

6 Contact Information

The SpamSieve Web site is located at http://www.c-command.com/spamsieve/. Questions about SpamSieve may be sent to mailto:support@c-command.com. I'm always looking to improve SpamSieve, so please feel free to send any feature requests to that address.

To make sure that you have the latest version of SpamSieve, you may wish to subscribe to the SpamSieve Announcements mailing list. The traffic on this list is very low, only one message per new version of SpamSieve. You may sign up using the form at http://www.c-command.com/spamsieve/support.shtml.

7 Registering

SpamSieve is shareware. If you find yourself using SpamSieve beyond a reasonable trial period, you must register it. Registration costs $10 (US) and entitles you to free updates and support.

To register, go to http://store.eSellerate.net/s.asp?s=STR804431608. Soon after paying, you'll receive an e-mail with your serial number. Enter it in the Registration window to personalize your copy of SpamSieve.

This is the honor system. If you use SpamSieve without registering, I probably won't know. However, registering will give me an incentive to continue updating and enhancing SpamSieve, and to write more Mac software. And you won't have to look at the "Unregistered" window anymore.

8 Version History

1.0—September 10, 2002

- First public release.

9 Legal Stuff

Please distribute the unmodified `SpamSieve-1.0.dmg` file on the Web, LANs, compilation CD-ROMs, etc. Please do not charge for it (beyond a reasonable cost for media), or distribute the contents of the image file in isolation.

This software is provided by the copyright holders and contributors "as is" and any express or implied warranties, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose are disclaimed. In no event shall the regents or contributors be liable for any direct, indirect, incidental, special, exemplary, or consequential damages (including, but not limited to, procurement of substitute goods or services; loss of use, data, or profits; or business interruption) however caused and on any theory of liability, whether in contract, strict liability, or tort (including negligence or otherwise) arising in any way out of the use of this software, even if advised of the possibility of such damage.

SpamSieve is a trademark of Michael Tsai. Mailsmith is a trademark of Bare Bones Software, Inc. Entourage is a trademark of Microsoft Corporation. Mac is a registered trademark of Apple Computer. All other products mentioned are trademarks of their respective owners.