SpamSieve 2.1.3 Manual

Michael Tsai
c-command.com

April 6, 2004

Contents

1 Introduction

1.1 What Is SpamSieve?

SpamSieve is an application that filters 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 them 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.

SpamSieve gives you back your inbox by bringing powerful Bayesian spam filtering to popular e-mail clients. It learns what your spam looks like, so it can block nearly all of it. It looks at your address book and learns what your good messages look like, so it won't confuse them with spam. Other spam filters get worse over time as spammers adapt to their rules; SpamSieve actually gets better over time as you train it with more messages. SpamSieve doesn't delete any messages—it only marks them in your e-mail client—so you'll never lose any mail. SpamSieve works with any number of mail accounts, of whatever types are supported by your e-mail software (e.g. POP, IMAP, Hotmail, AOL).

1.2 Identifying Spam

SpamSieve uses a statistical technique known as *Bayesian analysis*. For a more in-depth treatment of this technique applied spam, see this article by Paul Graham[1] and the papers it references. Bayesian spam filtering is highly accurate and adapts to new types of spam messages "in the field."

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 mean 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. (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.)

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

[1] `http://www.paulgraham.com/spam.html`

other spam filters become less effective as spammers figure out their rules, *are* because it has a larger corpus of your messages to work from.

1.3 Integration

Separate from the issue of identifying spam messages is the issue of how to prevent you from having to deal with them. There are basically six kinds of anti-spam software for doing this:

Challenge-Response Systems

This software requires people who sends you mail to prove that they are human, and not an automated spam-sending program. After sending you a message, they get a reply asking them to complete a task that is easy for humans but hard for computers. Only then is the message passed on to you. This system is a nuisance for senders, delays your reception of the mail, and becomes impractical when sending messages to a group of people. Also, challenge response systems cannot deal with spoofed senders or legitimate messages *are* sent by programs.

Server-Side Filters

This software runs on mail servers often filters out spam before you ever see it. This means that you do not have to download the spam messages that it catches. However, some spam messages may still get through, and, unless the filter is perfect, a few legitimate messages will not. These could be important messages, and you will never know that you lost them.

Server-Side Taggers

This variant of server-side filters does not delete messages before you download them. Instead, you download every message and configure your e-mail client to move messages that were tagged by the filter into a separate spam folder. This eliminates the major disadvantage of server-side filters—lost messages—however this type of filter is generally not as accurate as the ones below, because it does not adapt to your own mail.

Client-Side Filters

This software connects to your mail server to delete spam messages before your e-mail client can download them. This is a clunky approach: 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. You will also download every good message twice. 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 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.

Client-Side Proxies

This is like a client-side filter except that the proxy downloads messages once and stores them locally. The e-mail client then "downloads" the good messages from the proxy. This addresses the timing and double-download problems of client-side filters, but interaction with the filter is still awkward because it happens outside your e-mail client. In addition, you lose some control over connections to the mail server and which messages are left on the server.

Client-Side Integrated

This category includes SpamSieve and Apple Mail's built-in spam filter. Suspected spam messages are moved to a separate folder, which you can quickly scan at your leisure to make sure there are no false positives. The e-mail client downloads messages directly from the mail server, thus avoiding the problems of client-side filters and proxies. You can train the anti-spam software to improve its accuracy from inside your e-mail client, and accuracy is higher than with server-side filters because the anti-spam software can learn from the messages that *you* receive. You can also control how the spam filter interacts with your regular mail sorting rules.

1.4 Main Features

- Powerful Bayesian spam filtering results in high accuracy and almost no false positives. It adapts to the mail that *you* receive to get even better with time. Some other e-mail clients include Bayesian filters, but SpamSieve is more accurate.
- Integrates with your e-mail client for a superior user experience.
- Integrates with the Mac OS X Address Book (and also Eudora and Entourage's address books) so that messages from friends and colleagues are never marked as spam.
- Automatically maintains a blocklist so that it can instantly adapt to spam messages sent from particular addresses, and catch 100% of them.
- Automatically maintains a whitelist to guarantee that messages from particular senders or mailing lists are never marked as spam, without cluttering your address book with these addresses.
- You can customize the whitelist and blocklist, adding sophisticated rules that match various message headers, or the message body. The rules can match text in a variety of ways, including using regular expressions.

- Can honor Habeas[2] headers warranting that a message is not spam, as well as the "ADV" subject tag indicating that a message *is* spam.
- Many spammers encode the contents of their messages so that filters cannot see the incriminating words they contain. SpamSieve can decode and look inside these messages. Optionally it can mark them all as spam, on the theory that legitimate senders do not try to obscure their messages.
- SpamSieve keeps track of how accurate it is, how many good and spam messages you receive, and how these numbers change over time.
- Turn off new-mail notification in your e-mail client, and let SpamSieve notify you only when you receive non-spam messages.
- The corpus window and log let you see how each spam message was caught.

1.5 Why Choose SpamSieve?

Given that e-mail clients such as Apple Mail, Entourage, and Eudora include their own integrated spam filters, you may be wondering why you should consider SpamSieve. The answer is simple: SpamSieve's higher accuracy will save you time. Let's see how the experts compare it with these other program's filters:

Apple Mail

NetNewsWire creator Brent Simmons recently switched[3] to SpamSieve:

> Simply put: it catches my spam far more accurately than Mail ever did. Mail never came close.

Entourage

Macworld's April 2003 cover story, describing Entourage's filter:

> Poor; identified only 18 percent of spam; flagged 13 percent of legitimate mail as spam.

Macworld later honored[4] SpamSieve with a 2003 Editor's Choice Award and named it Software of the Year[5] (February 2004).

[2] `http://www.habeas.com`
[3] `http://inessential.com/?comments=1&postid=2764`
[4] `http://www.macworld.com/2003/12/news/eddys2003announce/`
[5] `http://www.macworld.com/2004/02/features/editorschoiceawards2004/`

Eudora

Macworld's review[6] of Eudora 6:

> I found that SpamSieve was a more effective spam blocker than Eudora 6's built-in filters.

In addition, SpamSieve works with the free Sponsored edition of Eudora; Eudora's own spam filter requires the $50 Paid version of Eudora.

SpamAssassin

Daring Fireball's John Gruber evaluated[7] SpamAssassin, which is installed on his ISP's mail server:

> So far in September, SpamSieve has been 99.7 percent accurate for me. About 3300 messages total, 2000 of which were spam. I've had 10 false negatives, and zero false positives. In fact, I haven't had a single false positive, ever, with any of the SpamSieve 2.0 betas.

This compares very favorably to SpamAssassin. Over the same period, SpamAssassin had over 90 false negatives—all of which SpamSieve caught.

2 Installing and Updating

2.1 Requirements

SpamSieve has been developed and tested on Mac OS X 10.2.6 and 10.3.3. It is designed to work with the following e-mail clients:

- Apple Mail[8] from Mac OS X 10.2.6 and later (10.3.x recommended)
- Emailer 2.0v3[9], previously available from Claris
- Entourage 9.0.1[10] and later (10.1.4 recommended) from Microsoft
- Eudora 5.2[11] and later (6.1 recommended) from Qualcomm

[6] `http://www.macworld.com/2004/01/reviews/eudora6/`
[7] `http://daringfireball.net/2003/09/interview_michael_tsai`
[8] `http://www.apple.com/macosx/features/mail/`
[9] `http://www.fogcity.com`
[10] `http://www.microsoft.com/mac/products/entouragex/entouragex.aspx?pid=entouragex`
[11] `http://www.eudora.com/mac`

- Mailsmith 1.5[12] and later (2.1.1 recommended) from Bare Bones Software
- Outlook Express 5.0[13] and later from Microsoft
- PowerMail 4.0[14] and later from CTM Development (5.0 recommended)

2.2 Updating From a Previous Version

SpamSieve will automatically read the corpus and statistics from previous versions. To update to the latest version of SpamSieve, you can simply quit the SpamSieve application and then replace the old application file with the new one. That is, if SpamSieve is installed in the `Applications` folder, drag the new SpamSieve application icon into the `Applications` folder and click **Replace** when the Finder asks if you want to overwrite the old version.

Apple Mail users should launch SpamSieve and choose **Install Apple Mail Plug-In and Scripts** from the **SpamSieve** menu.

Eudora users should launch SpamSieve and choose **Install Eudora Plug-In** from the **SpamSieve** menu.

2.3 Installing SpamSieve

Double-click the `SpamSieve-2.1.3.dmg` file to mount the SpamSieve disk image. Then move the SpamSieve application to your `Applications` folder.

Next, you *must* follow the instructions for setting up SpamSieve to use it with your e-mail client. After setting up SpamSieve, you will need to train it with examples of your spam messages and good messages.

2.4 Uninstalling SpamSieve

To uninstall SpamSieve, delete any rules that you created for it in your e-mail client. You can also delete the AppleScripts and/or plug-in that you installed. If you are using the SpamSieve Eudora Helper, delete it, and also run the Uninstall Eudora Helper program that came with SpamSieve. You can also delete the SpamSieve application and its data files,

[12] `http://www.barebones.com/products/mailsmith.html`
[13] `http://www.microsoft.com/mac/otherproducts/outlookexpress/outlookexpress.aspx?pid=outlookexpress`
[14] `http://www.ctmdev.com/powermail4.shtml`

which are stored in `~/Library/Application Support/SpamSieve`. The preferences file is stored in `~/Library/Preferences/com.c-command.SpamSieve.plist`.

3 Using SpamSieve With Your E-Mail Client

Before you can use SpamSieve, you must give it some examples of messages you consider to be spam, and ones which you do not. You do this by selecting some messages and then telling SpamSieve to add them to its corpus. For the details of how to do this, find the section below that corresponds to your e-mail client. For now, what's important is that you will train SpamSieve with both good messages and spam messages.

In general, the more messages you train SpamSieve with, the better its accuracy will be. For best results, you should train it with *at least* 600 messages. Training SpamSieve with more of one type of message will bias its predictions to that type. For instance, a corpus with mostly good messages will make SpamSieve conservative about identifying spam, leading to more false negatives. Most users find that *filling the corpus with about 65% spam messages* (as shown at the bottom of the **Statistics** window) produces a low level of false negatives, while keeping false positives rare or non-existent. For this reason, you should *not* add all your good messages to the corpus when you first install SpamSieve, because you likely do not have enough saved spam messages to compensate.

If SpamSieve marks a good message as spam, you should add the message to SpamSieve as a good message. This lets SpamSieve know that it made a mistake, and also adds the message to the corpus to improve future accuracy. Likewise, if SpamSieve marks a spam message as good, you should add the message to SpamSieve as a spam message. *If you do not correct SpamSieve when it makes mistakes, its accuracy will deteriorate over time.*

If you make a mistake and tell SpamSieve that a message is spam when it is actually good (or vice-versa), simply correct yourself as you would correct SpamSieve. That is, if the message is good, add the message as good; if it is spam, add it as spam. SpamSieve will "undo" the previous, incorrect, addition to the corpus.

To improve SpamSieve's accuracy, it is important to train it with new messages as you receive them. When they first install SpamSieve, most users have many good messages to train it with, but few spams. For this reason, SpamSieve is set to automatically train itself with spam messages as you receive them. It does not automatically train itself with good messages, on the assumption that you have already trained it with plenty of good mail. Both of these settings may be modified in the preferences. As a result, after training SpamSieve for the first time, you only need to train it to correct mistakes.

3.1 Apple Mail

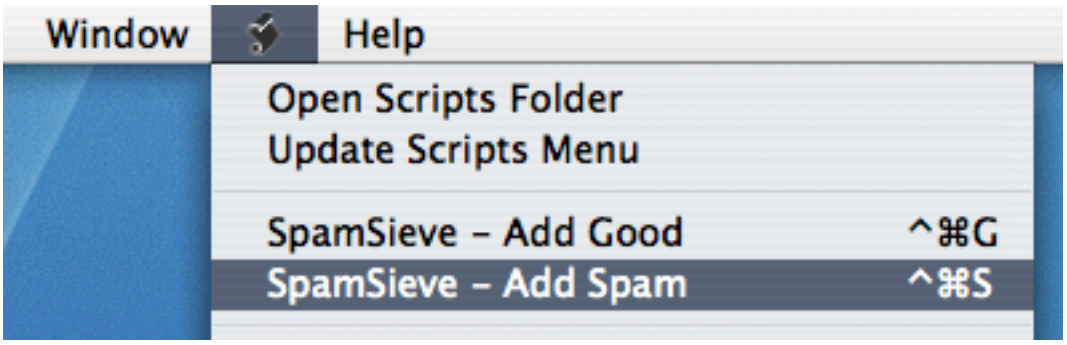
3.1.1 Installing on Mac OS X 10.3.x

Follow these instructions if you are using Mac OS X 10.3.x. If you have an older version of Mac OS X, please see the next section.

First, go to the **Junk Mail** tab of Apple Mail's **Preferences** window and uncheck **Enable Junk Mail filtering**. This will disable Apple Mail's junk mail filter so that it does not interfere with SpamSieve.

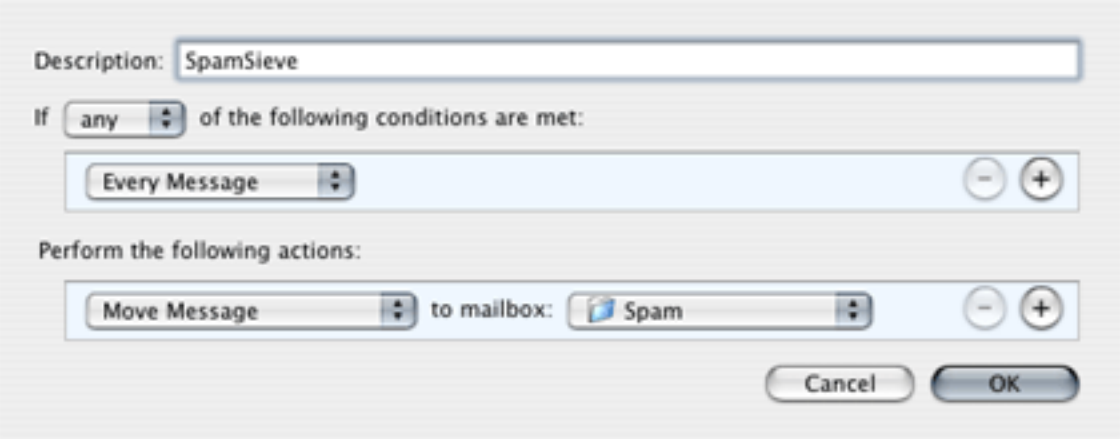
Choose **New** from Apple Mail's **Mailbox** menu, and create a new mailbox called **Spam** with location **On My Mac**.

Double-click the SpamSieve application and choose **Install Apple Mail Plug-In and Scripts** from the **SpamSieve** menu. (If you are using a preferences-swapping utility such as Location X[15], you will need to do that for each location.) When you open Mail, you should see the two SpamSieve items in the **Scripts** menu:



Choose **Preferences**... from the **Mail** menu and click on **Rules**. Click the **Add Rule** button. Change the description to **SpamSieve**. (The description of the rule *must* start with **SpamSieve**.) Change the **From** menu to say **Every Message**. (Or, you can choose a more restrictive criterion, if you want.) Then, next to **Move Message**, select the **Spam** folder you just created. The rule should now look like:

[15] `http://homepage.mac.com/locationmanager/`



and you can click **OK** to close it and save your changes. Please note that although the rule *looks* like it will move every message to the **Spam** folder, because you have installed SpamSieve's plug-in, it will only move the spam messages.

SpamSieve will now automatically move new spam messages that you receive to a folder called **Spam**. If SpamSieve is not running when you receive new messages, it will launch automatically.

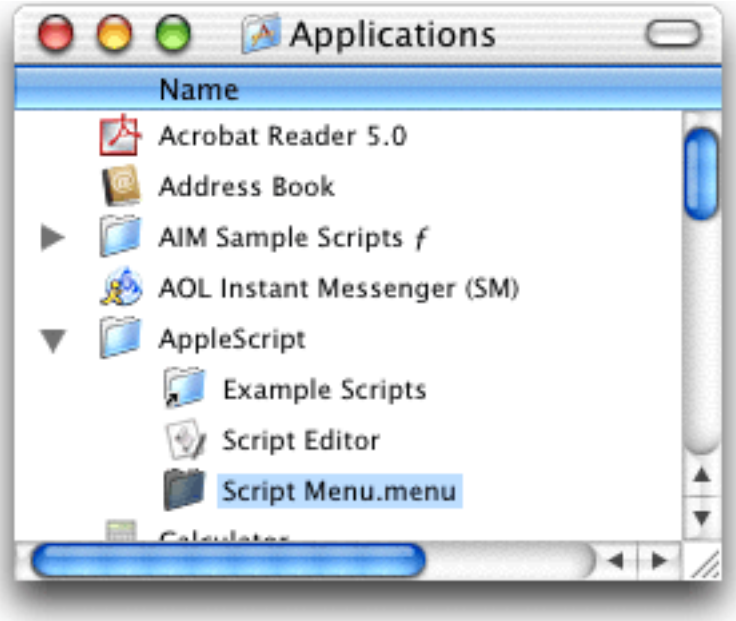
3.1.2 Installing on Mac OS X 10.2.x

Follow these instructions if you are using Mac OS X 10.2.x. If you have a later version of Mac OS X, please see the previous section.

First, go to the **Junk Mail** submenu of the **Mail** menu, and choose **Off**. This will disable Apple Mail's junk mail filter so that it does not interfere with SpamSieve.

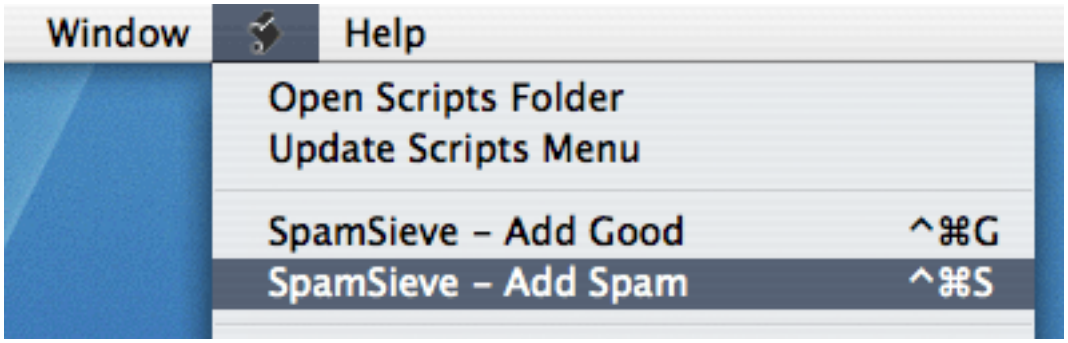
Next, locate the `Script Menu.menu` file in the `AppleScript` folder inside the `Applications` folder. `Script Menu.menu` may look like a folder. Most Mac OS X 10.2 users already have this file installed. Unfortunately, Apple no longer provides it as a seprate download. If you do not have it, you can update to Mac OS X 10.3 (in which Mail has a script menu built-in) or use a third-party script menu such as FastScripts[16].

[16] `http://www.red-sweater.com/RedSweater/FastScripts.html`



Double-click `Script Menu.menu`; this will add a menu to the right side of the menu bar, with the icon of a script. You only need to do this step once; Mac OS X will remember to show the menu the next time you log in.

Double-click the SpamSieve application and choose **Install Apple Mail Plug-In and Scripts** from the **SpamSieve** menu. (If you are using a preferences-swapping utility such as Location X[17], you will need to do that for each location.) When you open Mail, you should see the two SpamSieve items in the **Scripts** menu:

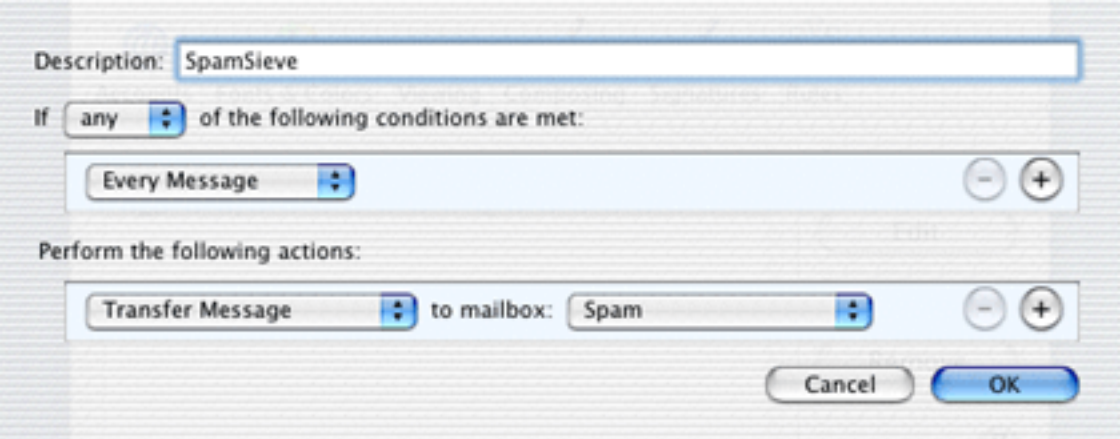


Choose **New**... from the **Mailbox** menu and create a new mailbox named **Spam**.

Choose **Preferences**... from the **Mail** menu and click on **Rules**. Click the **Add Rule** button. Change the description to **SpamSieve**. (The description of the rule *must* start

[17] `http://homepage.mac.com/locationmanager/`

with **SpamSieve**.) Change the **From** menu to say **Every Message**. (Or, you can choose a more restrictive criterion, if you want.) Then, next to **Transfer Message**, select the **Spam** folder you just created. The rule should now look like:



and you can click **OK** to close it and save your changes. Please note that although the rule *looks* like it will move every message to the **Spam** folder, because you have installed SpamSieve's plug-in, it will only move the spam messages.

SpamSieve will now automatically move new spam messages that you receive to the **Spam** folder.

3.1.3 Training SpamSieve

To train SpamSieve with spam messages, select one or more of them and then choose **SpamSieve - Add Spam** from the **Scripts** menu. The messages will be colored in gray and moved to the **Spam** folder. To train SpamSieve with good messages, select one or more of them and then choose **SpamSieve - Add Good** from the **Scripts** menu.

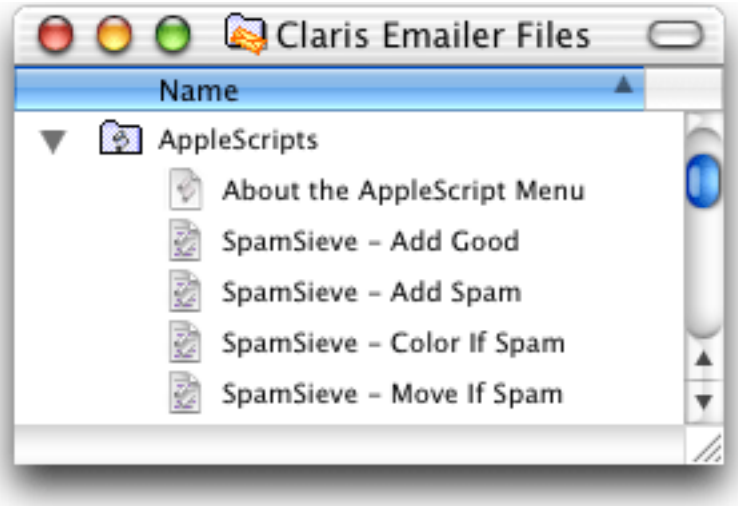
3.1.4 Manually Processing Messages

To manually ask SpamSieve to move messages that it thinks are spam to the **Spam** mailbox, select the messages and choose **Apply Rules** from the **Message** menu.

3.2 Emailer

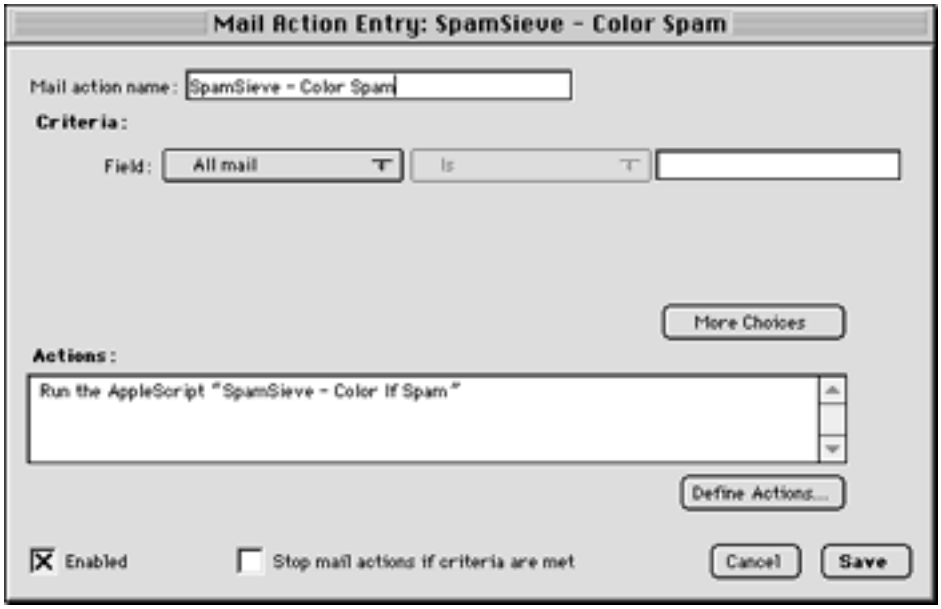
3.2.1 Installing

Choose **Show Other Scripts** from SpamSieve's **SpamSieve** menu. Copy the files from the `For Emailer Users` folder into Emailer's `AppleScripts` folder:



You may need to quit and re-launch Emailer in order for it to notice that you have installed the SpamSieve AppleScripts.

If you want SpamSieve to color messages that it thinks are spam, set up a mail action in Emailer that looks like this:



If, instead, you want SpamSieve to move suspected spam messages to a **Spam** folder (that it creates automatically), set up a mail action in Emailer that looks like this:

SpamSieve will now automatically color or move new spam messages that you receive, depending on which mail action you set up. If SpamSieve is not running when you receive new messages, it will launch automatically.

3.2.2 Training SpamSieve

To train SpamSieve with spam messages, select one or more of them and then choose **SpamSieve - Add Spam** from Emailer's **Scripts** menu. To train SpamSieve with good messages, select one or more of them and then choose **SpamSieve - Add Good** from Emailer's **Scripts** menu.

3.2.3 Manually Processing Messages

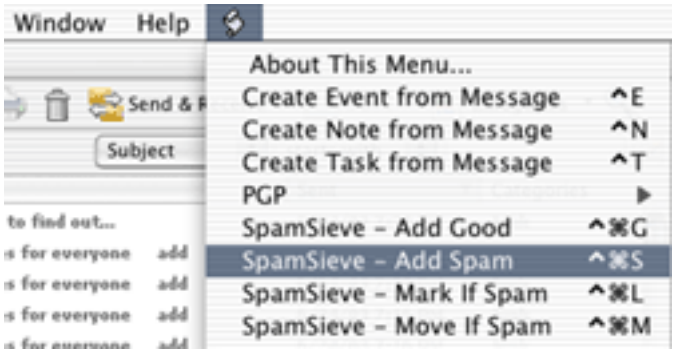
To manually ask SpamSieve to color or move messages that it thinks are spam, select the messages and choose **SpamSieve - Color If Spam** or **SpamSieve - Move If Spam** from Emailer's **Scripts** menu.

3.3 Entourage

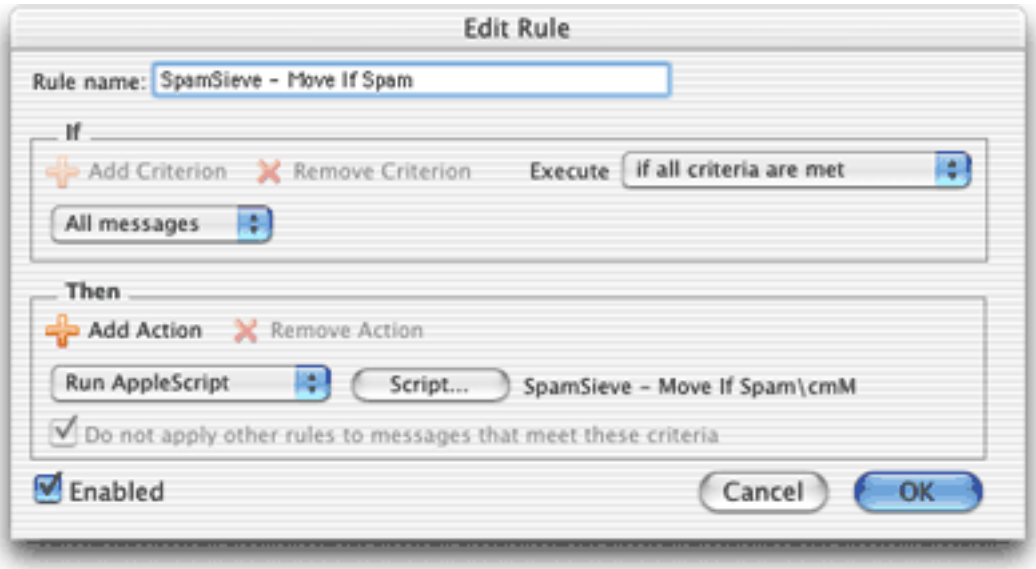
3.3.1 Installing

Go to Entourage's **Tools** menu select **Junk Mail Filter**. Make sure the Junk Mail Filter is disabled because it interferes with SpamSieve. Then open the **Mailing List Manager**, also in the **Tools** menu, and see if there are any items listed there. If so, you can leave them for now, but if you find that SpamSieve isn't processing certain messages, you should try removing the items from the Mailing List Manager.

Double-click the SpamSieve application and choose **Install Entourage Scripts** from the **SpamSieve** menu. After you quit and re-launch Entourage, you should see four SpamSieve items in Entourage's **Scripts** menu:



Now, set up a mail rule in Entourage that looks like this:



To do this, choose **Rules** from Entourage's **Tools** menu. Click on the tab corresponding to the type of account you have (e.g. POP). If you have more than one kind of account, you will need to create a rule for each account type. Click the **New** button. Type a name for your rule. Then click just to the left of **Change status** to select the first action. Click **Remove Action**. Click on the menu that says **Set category** and select **Run AppleScript**. Then click the **Script**... button and select the `SpamSieve - Move If Spam` file. The rule window should now look like the above screenshot. This will make SpamSieve move suspected spam messages to a **Spam** folder (that it creates automatically).

It is important that you create the rule exactly as shown. Do not add additional actions below the action that runs the AppleScript. Such actions would apply to all messages, which is probably not what you want. To customize what Entourage does when SpamSieve finds a spam message, you need to edit the AppleScript rather than the rule.

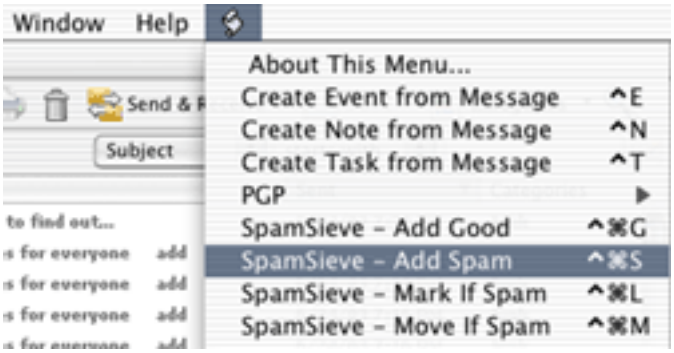
Finally, go to SpamSieve's **Preferences** window. Make sure that **Use Entourage address book** is checked, and click the **Load** button.

SpamSieve will now automatically move new spam messages that you receive. If SpamSieve is not running when you receive new messages, it will launch automatically.

3.3.2 Training SpamSieve

To train SpamSieve with spam messages, select one or more of them and then choose **SpamSieve - Add Spam** from Entourage's **Scripts** menu. To train SpamSieve with good

messages, select one or more of them and then choose **SpamSieve - Add Good** from Entourage's **Scripts** menu.



Note that due to the way Entourage's AppleScript interface works, you may not be able to train SpamSieve by selecting messages in custom mail views. Instead, select the messages in their actual folders.

Make sure that you correct SpamSieve's mistakes by using the commands in the **Scripts** menu—do not click the underlined blue text to indicate that a message is not spam.

3.3.3 Manually Processing Messages

To manually ask SpamSieve to mark or move messages that it thinks are spam, select the messages and choose **SpamSieve - Mark If Spam** or **SpamSieve - Move If Spam** from Entourage's **Scripts** menu.

3.3.4 IMAP Accounts

Entourage does not support moving IMAP messages via AppleScript, so if you use IMAP the **SpamSieve - Move If Spam** script will not move spam messages into your **Spam** folder.

The `SpamSieve - Add Spam` and `SpamSieve - Move If Spam` scripts contain an experimental workaround for moving IMAP messages. You can enable this by editing the scripts in Script Editor and changing the `false` after `tryToMoveIMAPMessages` to `true`. However, some users have found that this exposes a bug in Entourage, causing it to crash.

If you do not require IMAP, you can try creating a POP account in Entourage and re-entering your account information. Many IMAP accounts also work via POP, and this will

allow SpamSieve to move the messages that it thinks are spam.

Alternatively, you can create an Entourage rule that moves messages that SpamSieve has marked as junk into another folder. After receiving mail, manually apply this rule to the messages in your IMAP account.

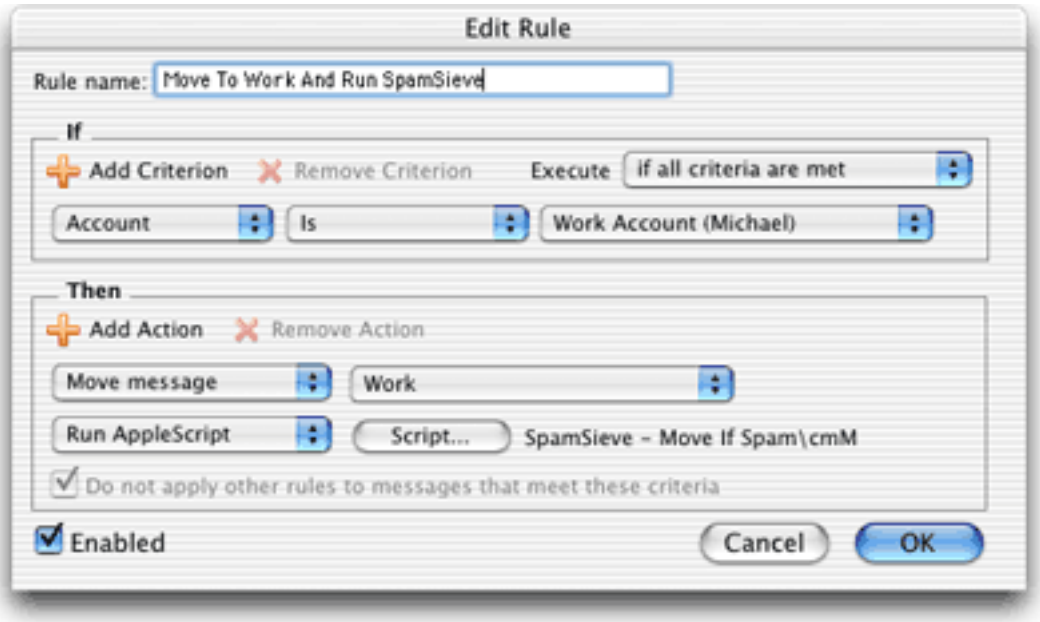
Yet another alternative is to use a custom view. Make a custom view of your IMAP account that looks for messages that are not junk. In this way, you can view your good messages without being distracted by spams.

3.3.5 Advanced Rules

Two complications are that once an Entourage rule runs an AppleScript or moves a message, it cannot apply any more rules to that message. Here are some ways to work around this.

One option is to order your rules so that Entourage applies the SpamSieve rule after all your other rules. You can change the order of the rules by choosing **Rules** from Entourage's **Tools** menu and dragging the rules in the list to change their order. With this approach, you can filter your good messages into folders however you want. Any mail that is not moved into another folder will remain in your inbox. Then, the SpamSieve rule will either mark the spams as junk or move them to a **Spam** folder. The disadvantage to this approach is that SpamSieve cannot catch any spams among the messages that were moved by your other rules.

Another option is to add the **Run AppleScript** action to each rule that moves messages. For instance, suppose you have a rule that moves all the messages from your **Work Account** account into a **Work** folder. You could set up the rule like this:



Now, messages sent to that account will be moved to the **Work** folder. Spam messages sent to that account will be moved to the **Spam** folder.

You can add the SpamSieve AppleScript action to every rule that moves messages and also to a "catch-all" rule that applies to messages that aren't moved. Then SpamSieve will be able to filter all the messages that you receive.

Please contact spamsieve@c-command.com[18] if you have trouble setting up Entourage to filter messages the way you want.

3.4 Eudora 6

3.4.1 Installing

If you are using Eudora 5.2, please see the Eudora 5.2 section.

Double-click the SpamSieve application and choose **Install Eudora Plug-In and Scripts** from the **SpamSieve** menu. SpamSieve will install its plug-in and reveal the Eudora application file for you.

Select the Eudora application file and choose **Get Info** from the Finder's **File** menu. Expand the **Plug-ins** pane and click **Add**.... Check the **Esoteric Settings** plug-in to enable it.

[18] `mailto:spamsieve@c-command.com`

Make sure that **SpamWatch OSX** and **SpamHeaders OSX** are unchecked. You can now close the info window and launch Eudora.

When you start up Eudora, you should see SpamSieve listed in the **About Message Plug-ins**... window that is accessible from the **Eudora** menu. Choose **Preferences** from the **Eudora** menu, scroll down to the **Junk Extras** settings panel (which is at the very bottom), and check **Always enable Junk/Not Junk menu items**. Next, select the **Junk Mail** pane and make sure that **Hold junk in Junk mailbox** is checked. If you are using IMAP, make sure that **Run junk scoring plugins on this IMAP account** is checked in the **IMAP** settings pane.

Now Eudora will use SpamSieve to filter all incoming messages. It will move the spam messages to the **Junk** mailbox. This happens *before* Eudora runs any incoming message filters that you have set up.

Normally, Eudora will launch the SpamSieve application when new messages arrive or when you train SpamSieve from inside Eudora. However, on some machines, it will not launch SpamSieve automatically. In this case, you must manually open the SpamSieve application when you want Eudora to filter spam messages.

3.4.2 Training SpamSieve

To train SpamSieve with spam messages, select one or more of them and then choose **Junk** from Eudora's **Message** menu. To train SpamSieve with good messages, select one or more of them and then choose **Not Junk** from Eudora's **Message** menu.

3.4.3 Setting Options

Eudora applies SpamSieve to all incoming messages. The **Junk Mail** area of Eudora's preferences lets you customize how Eudora interacts with SpamSieve. Note that the **Junk Threshhold** [**sic**] slider will have no effect because SpamSieve always gives Eudora scores that are exactly 0 or exactly 100. Instead of using this slider, you should use the one in the **Bayesian** tab of SpamSieve's preferences.

Other settings in the **Junk Mail** pane do affect SpamSieve. For instance, if you check **Mail isn't junk if the sender is in an address book**, then Eudora will not pass those messages along to SpamSieve; it will assume that they are good. Note that Eudora always considers your address to be in its address book, even though it may not be explicitly listed there. Thus, if you receive spam that is forged so that it appears to be sent from your own address, you must uncheck **Mail isn't junk if the sender is in an address book** in order

for SpamSieve to catch it. (To still have SpamSieve whitelist your address book, you can export[19] the Eudora address book to vCard and then import it into the Mac OS X Address Book).

The **Junk Extras** area of Eudora's preferences lets you control some additional settings, such as whether junk messages are removed from the mail server.

3.5 Eudora 5.2

3.5.1 Installing

Using SpamSieve with Eudora 6 is highly recommended. However, SpamSieve can also work with Eudora 5.2, and some Eudora 6 users may prefer the configuration described here because it is more customizable.

Choose **Show Other Scripts** from SpamSieve's **SpamSieve** menu. Move the `SpamSieve Eudora Helper` file in the `For Eudora 5.2 Users` folder to the `Applications` folder of your hard disk. You will need to launch this applet the first time you use SpamSieve with Eudora.

Also in the `For Eudora 5.2 Users` folder is the `Uninstall Eudora Helper` file. Run this applet if you no longer want to use SpamSieve with Eudora.

Create a mailbox in Eudora called **Spam** that is at the same level as the **In** mailbox. When you receive new spam messages, SpamSieve will move them to the **Spam** mailbox. It will also mark good messages by setting their priority to lowest (indicated by two downward pointing carets) and mark spam messages by setting their status to transfer error (indicated by a red "X").

3.5.2 Training SpamSieve

To train SpamSieve with spam messages, select one or more of them. Choose the **Filter Messages** command in Eudora's **Special** menu. Then double-click **Add Spam**.

[19] `http://homepage.mac.com/aamann/Eudora_vCard_Export.html`

To train SpamSieve with good messages, select one or more of them. Choose the **Filter Messages** command from Eudora's **Special** menu. Then double-click **Add Good**.

3.5.3 Manually Processing Messages

To manually ask SpamSieve to mark or move messages that it thinks are spam, select one or more of them. Choose the **Filter Messages** command in Eudora's **Special** menu. Then double-click **Filter**.

3.5.4 Setting Options

By configuring the `SpamSieve Eudora Helper` applet, you can tell SpamSieve to process your good messages and spam messages in other ways. First, quit the applet. Then open it using the Script Editor program in the `AppleScript` folder of your `Applications` folder. The top of the script contains the following lines:

```
property moveToSpamFolder     : true  -- moves spams to a "Spam" mailbox
property markSpamMessages     : true  -- marks spams with red x
property markSpamMessagesRead : false -- marks spams as "already read"
property labelSpamMessages    : false -- colors spam messages brown
property markGoodMessages     : true  -- marks good messages with carets
property labelGoodMessages    : false -- colors good messages green
property removeSpamMessagesFromServer : false
```

You can change a `false` to `true` or a `true` to `false` to set the options the way you want. For instance, to have SpamSieve not move spam messages into a separate mailbox, change the `true` in the first line to `false`. When you are finished making changes, choose **Save**

in Script Editor's **File** menu, close the window, and then re-launch the SpamSieve Eudora Helper.

3.5.5 Eudora Limitations

The following limitations are due to problems with Eudora's "notification" interface. Because of these limitations it is recommended that you use Eudora 6 and the `SpamSieve Eudora Plug-In`, as described in the Eudora 6 section. The plug-in avoids these limitations.

- Eudora gives messages to SpamSieve *after* all the other filters have run. It is not possible to change this ordering.
- SpamSieve cannot filter messages that are moved by other filters. For instance, if you have a filter that moves incoming messages from Steve Jobs to a separate mailbox, SpamSieve will not mark any of those messages as spam, even if a spammer pretends to be Jobs. This limitation applies to both automatic filtering of incoming mail and manual filtering of selected messages.
- Sometimes the wrong message is marked. That is, SpamSieve may decide that message A is spam and ask Eudora to mark it with a red "X"; in rare circumstances, Eudora will instead mark some other message B with the "X." You can tell if this has happened by comparing SpamSieve's log to the way the messages are marked in Eudora. This problem seems to occur when the **In** mailbox sorted.
- Sometimes SpamSieve never sees a message that should have been filtered. You can tell if this has happened by the absence of that message in the log. It may help to remove any "notify user" filter action that you have set up.
- Sometimes SpamSieve determines that a message is good or spam, but Eudora does not mark it all. You can tell if this has happened by comparing SpamSieve's log to the way the messages are marked in Eudora.
- SpamSieve cannot add or filter messages that are stored in the Trash mailbox or in mailbox files outside the `Mail Folder` folder in the `Eudora Folder`. Note that this includes all IMAP messages. To access these messages, first move them to a non-trash mailbox file that is stored inside the `Mail Folder` folder.
- If you manually apply filters while Eudora is in the process of downloading mail, Eudora will show the SpamSieve dialog box twice. If this happens, just choose **Skip** the second time.
- Sometimes Eudora erroneously shows the SpamSieve dialog when you check for new mail.

- Sometimes after a long delay in talking to the mail server, Eudora stops notifying SpamSieve when it receives new messages. You can work around this by quitting and re-launching the SpamSieve Eudora Helper.

3.6 Mailsmith

3.6.1 Installing

Mailsmith 2.0 and later feature direct integration with SpamSieve. This is more convenient and easier to use than the script- and filter-based integration that was necessary when using previous versions of Mailsmith. You can enable SpamSieve simply by clicking the checkbox in the **Spam Handling** pane of Mailsmith's preferences. For more information about using SpamSieve with Mailsmith, please see Chapter 8 of the Mailsmith User Manual.

3.6.2 Training SpamSieve

To train SpamSieve with spam messages, select one or more of them and then choose **Mark as Spam** from Mailsmith's **Message** menu. To train SpamSieve with good messages, select one or more of them and then choose **Mark as Non-Spam** from Mailsmith's **Message** menu.

3.6.3 Setting Options

You can configure how Mailsmith and SpamSieve work together from the **Spam Handling** pane of Mailsmith's preferences. Checking the **Train the Spam Detector** checkbox here is equivalent to checking *both* auto-training checkboxes in SpamSieve's preferences. Unchecking **Train the Spam Detector** causes SpamSieve to use the settings in its **Preferences** window. It is recommended that you uncheck **Train the Spam Detector** once you have trained SpamSieve with a few hundred messages of each type.

3.6.4 Identifying Spam Messages

When using SpamSieve with Mailsmith, Mailsmith tags messages using the **Is Spam** and **Is Not Spam** properties. Although you can use Mailsmith's **Advanced Query** feature to search on these properties, they are otherwise not visible in the user interface. Therefore, you should mark spam messages in a visible way, either by letting Mailsmith move them to

a separate mailbox, or by setting up a filter to change the messages' labels based on their **Is Spam** and **Is Not Spam** properties. Otherwise, you will not be able to correct SpamSieve's mistakes to improve its accuracy.

3.6.5 Mailsmith Extras

Choose **Show Other Scripts** from SpamSieve's **SpamSieve** menu. This reveals the `Mailsmith Extras` folder, which contains AppleScripts for use with Mailsmith. These make it possible for scripters to further customize and automate the labelling and marking of messages in Mailsmith. If you are using Mailsmith 2.0 and do not write AppleScripts, you can ignore the `Mailsmith Extras` folder.

You can add the AppleScripts to Mailsmith's **Scripts** menu by copying them to the `Scripts` folder inside the `Mailsmith Support` folder. A filter such as the following may be used to change the labels of incoming spam messages.

This filter will pass all messages along to SpamSieve for analysis. It will set the **Is Spam** or **Is Not Spam** property of the message, and change the label of the message if it is spam. This is roughly equivalent to enabling SpamSieve in Mailsmith's preferences, but because it uses AppleScript it is more customizable. Additionally, you can change the filter criteria to pass only select messages along to SpamSieve.

For best results, use either Mailsmith's direct integration with SpamSieve or AppleScripts like those in the `Mailsmith Extras` folder. Do not mix and match them.

3.7 Outlook Express

3.7.1 Installing

Choose **Show Other Scripts** from SpamSieve's **SpamSieve** menu. Copy the files from the `For Outlook Express 5 Users` folder to the `Script Menu Items` folder inside the `Outlook Express 5.0.3 folder` folder.

Follow the instructions for Entourage to create a rule in Outlook Express that applies the `SpamSieve - Move If Spam` script. SpamSieve will now move new spam messages that you receive to the **Spam** folder. If SpamSieve is not running when you receive new messages, it will launch automatically.

3.7.2 Training SpamSieve

To train SpamSieve with spam messages, select one or more of them and then choose **SpamSieve - Add Spam** from Outlook Express's **Scripts** menu. To train SpamSieve with good messages, select one or more of them and then choose **SpamSieve - Add Good** from Outlook Express's **Scripts** menu.

Make sure that you correct SpamSieve's mistakes by using the commands in the **Scripts** menu—do not click the underlined blue text to indicate that a message is not spam.

3.7.3 Manually Processing Messages

To manually ask SpamSieve to move messages that it thinks are spam, select the messages and choose **SpamSieve - Move If Spam** from Outlook Express's **Scripts** menu.

3.8 PowerMail 5

3.8.1 Installing

If you are using PowerMail 4, please see the PowerMail 4 section.

PowerMail 5 ships with built-in support for SpamSieve. To enable it, go to the **Spam** pane of PowerMail's **Preferences** window. Choose **SpamSieve** from the **Third party spam**

filter pop-up menu. For more information about how to configure PowerMail's handling of spam messages, please see the PowerMail documentation.

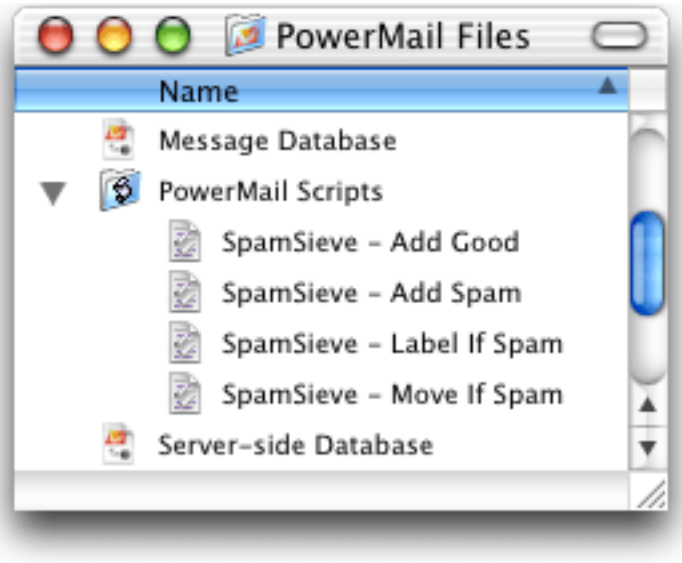
3.8.2 Training SpamSieve

To train SpamSieve with spam messages, select one or more of them and then choose **Mark as Spam** from the **Mail** menu. To train SpamSieve with good messages, select one or more of them and then choose **Mark as Not Spam** from the **Mail** menu.

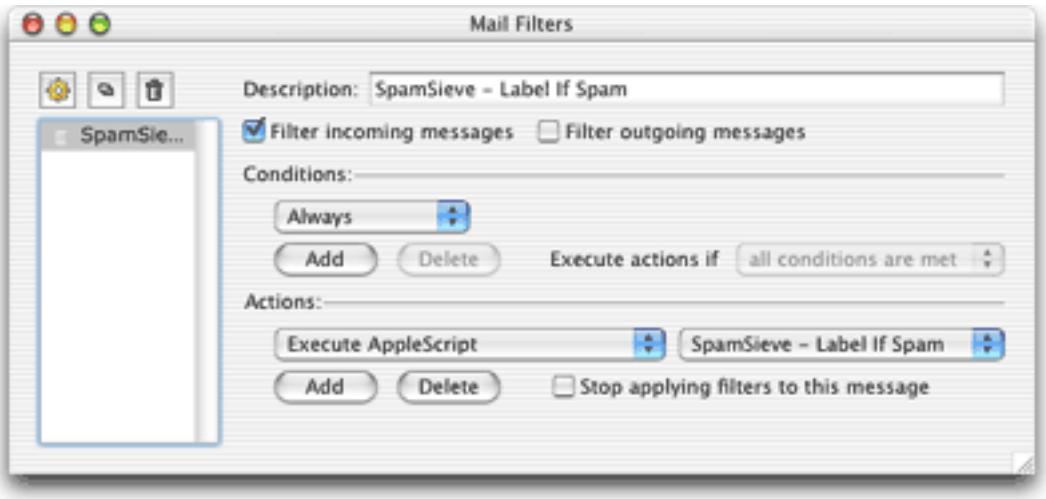
3.9 PowerMail 4

3.9.1 Installing

Choose **Show Other Scripts** from SpamSieve's **SpamSieve** menu. Copy the files from the `For PowerMail 4.x Users` folder to the `PowerMail Scripts` folder inside the `PowerMail Files` folder. The `PowerMail Files` folder is probably located in your `Documents` folder.



If you want SpamSieve to label messages that it thinks are spam, set up a filter in PowerMail that looks like this:



If, instead, you want SpamSieve to move suspected spam messages to a **Spam** folder (that it creates automatically), set up the filter to use the `SpamSieve - Move If Spam` script instead.

SpamSieve will now mark or move new spam messages that you receive. If SpamSieve is not running when you receive new messages, it will launch automatically.

3.9.2 Training SpamSieve

To train SpamSieve with spam messages, select one or more of them and then choose **SpamSieve - Add Spam** from PowerMail's **Scripts** menu. To train SpamSieve with good messages, select one or more of them and then choose **SpamSieve - Add Good** from PowerMail's **Scripts** menu.

3.9.3 Manually Processing Messages

To manually ask SpamSieve to label or move messages that it thinks are spam, select the messages and choose **SpamSieve - Label If Spam** or **SpamSieve - Move If Spam** from PowerMail's **Scripts** menu.

3.9.4 IMAP Accounts

PowerMail does not support moving IMAP messages via AppleScript, so if you use IMAP the **SpamSieve - Move If Spam** script will not move spam messages into your **Spam**

folder. IMAP users should create the rule that uses the labelling script.

4 E-Mail Client Customization

SpamSieve works with your e-mail client to filter out spam messages. First, the mail client downloads new messages from the mail server. The mail client passes the messages to a plug-in or AppleScript, which in turn passes them to the SpamSieve application. SpamSieve analyzes the messages to see whether they are spam. It returns the verdict to the plug-in or script, which then directs the mail client to move the messages to another folder.

4.1 Rule Ordering

Some of the e-mail clients that SpamSieve supports let you control the order in which the rules (a.k.a. filters or mail actions) that you have created process mail. How you order the SpamSieve rule is up to you. If you get a lot of spam that matches the rules you use to organize your mail, you might want to run the SpamSieve rule first. This will allow it to find spam among all your messages. If you would 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 and file away messages from known senders. Be sure to check the SpamSieve preferences for additional filtering options.

4.2 Moving Spam Messages After Training

Normally when you add spam messages to the corpus, SpamSieve moves the messages to a **Spam** folder in your e-mail client. If you prefer that it not move the messages, you can do the following. Use Script Editor to open the `SpamSieve - Add Spam` AppleScript for your e-mail client. (Mailsmith and Eudora users do not have such a script.) The first line of the script contains the text:

```
property shouldMoveSpam : true
```

Change the `true` to `false` so that you have:

```
property shouldMoveSpam : false
```

Then save the script and close its window.

There are other aspects of the SpamSieve AppleScripts that you can easily customize by editing the scripts in Script Editor. Take a look at the different properties listed at the top of each script.

4.3 Compatibility Notes

Some Mac e-mail clients only work with the older resource-fork format for compiled AppleScripts. This format is no longer the Script Editor default in Mac OS X 10.2 and later. To save scripts in the older format, make sure that you edit the supplied scripts (or copies of them) rather than creating new compiled script files from inside Script Editor. If you create a new file, Script Editor will save it in the, incompatible format.

Script Editor 2.0 cannot edit the `SpamSieve Eudora Helper` applet. To customize the applet, use Script Editor 1.9.

4.4 Integrating With Other Applications

SpamSieve's interface for integrating with third-party mail and news programs is completely open. It is possible to add support for additional programs simply by writing some AppleScripts. SpamSieve's AppleScript dictionary contains some basic information about the supported commands. However, there are some subtle, but important, points that are not discussed in the dictionary's documentation. If you would like to connect an application to SpamSieve, please contact spamsieve@c-command.com[20] so that I may assist you.

5 Menus

5.1 The File Menu

5.1.1 Import Corpus...

This imports the words in a corpus that was exported in XML format. This is the same format used by SpamSieve 1.x. Importing a corpus merges it with the active corpus. To

[20] `mailto:spamsieve@c-command.com`

replace the active corpus with the one you are importing, use the **Reset Corpus**... command before importing.

5.1.2 Export Corpus...

This exports the active corpus to XML format. You might do this in order to use the corpus on another machine.

5.1.3 Import Messages...

This imports messages stored in mbox format. You can select whether the mbox file contains good messages or spam messages. To replace the active corpus with the messages you are importing, use the **Reset Corpus**... command before importing.

5.1.4 Import Seed Spam...

Some users do not have many saved spam messages with which to train SpamSieve. This command adds about 1400 spam messages (from a public archive) to the corpus in order to jump-start spam recognition. You still need to train SpamSieve with your own good messages, however. Note that accuracy will be better if you train SpamSieve with your own spam rather than the seed spam. Thus, you should only import seed spam if you have few saved spam messages and do not receive many new ones per day.

5.2 The Filter Menu

5.2.1 Show Corpus

This command opens the **Corpus** window so that you can examine the words that SpamSieve has found in your e-mails. You can click on the name of a column to sort by that column. Click again on the column to reverse the sort direction. The meanings of the columns are as follows:

Word

A word in the corpus.

Spam

The number of times the word has occurred in spam messages.

Good

The number of times the word has occurred in good messages.

Total

The total number of times the word has occurred.

Prob.

The probability that a message is spam, given that it contains the word (and in the absence of other evidence).

Last Used

The date that the word was added to the corpus, or the date that it last appeared in a received message (whichever is later).

You can copy the selected rows to the clipboard or drag and drop them into another application.

With the window sorted by **Word**, you can type the first few letters of a word to locate that word in the corpus. Similarly, you can sort by one of the other columns and type a number to locate the first word whose value for the sorted column matches the number you typed.

You can edit the spam and good counts associated with a word by double-clicking on the number in the **Spam** or **Good** column. Changing the numbers for important words can greatly affect SpamSieve's accuracy, so you shouldn't make changes without good reason.

You can remove words that you don't want in the corpus by selecting them and pressing Delete.

5.2.2 Prune Corpus...

This command removes from the corpus words that have not been used in a set number of days. This can decrease SpamSieve's memory use and increase its speed. However, pruning can reduce SpamSieve's accuracy, so you should only prune if you find that SpamSieve is using too many system resources.

5.2.3 Reset Corpus...

This command removes all the words and messages from the corpus. This will enable you to retrain SpamSieve from scratch, and SpamSieve will let you use your old messages in the retraining.

5.2.4 Show Statistics

This command opens the **Statistics** window, which displays the following information:

Good Messages
: The number of non-spam messages that SpamSieve has filtered.

Spam Messages
: The number of spam messages that SpamSieve has filtered.

False Positives
: The number of good messages that SpamSieve identified as spam.

False Negatives
: The number of spam messages that SpamSieve identified as good.

% Correct
: The percent of messages that SpamSieve identified correctly.

Good Messages
: The number of non-spam messages that are used to identify spam messages.

Spam Messages
: The number of spam messages that are used to identify spam messages.

Total Words
: The total number of unique words in the corpus.

You can copy all the statistics to the clipboard using the **Copy** command in the **Edit** menu or by clicking the **Copy Stats** button.

Normally, SpamSieve shows statistics for all the messages that it processed since it was installed. Because the accuracy and the number of messages you receive change with time,

you may wish to only see recent statistics. Click the **Set Date**... button at the bottom of the window to control which old statistics are hidden from view. You can edit the date and time shown in the sheet, or enter an entirely new date. SpamSieve will accept dates specified in natural language, such as "last Sunday at dinner" or "today."

5.2.5 Open Log

SpamSieve keeps a log of messages that it has filtered, words that it has used to decide whether the messages were spam, messages you have added to the corpus, and any errors that have occurred. This command opens the log file so that you can look at it. Normally, there is no reason (aside from curiosity) to look at the log file. However, if you believe SpamSieve is not working as it should, the log file provides useful information about what SpamSieve has done. If you find that the log file is taking up too much disk space, you can delete it at any time. SpamSieve will then start a new log file as needed.

5.2.6 Show Blocklist

This opens the **Blocklist** window. The blocklist consists of a list of *rules*. If a message *matches* one or more rules on the blocklist, SpamSieve will predict that it is spam. Each row in the blocklist window represents one rule. The meanings of the columns are as follows:

Date

The date that the rule was added to the blocklist.

Header

The part of the message that will be matched against the rule.

Match Style

How the rule text will be matched against the text from the message's header.

Text to match

The text that will be matched against the message's header.

√

If this is checked, the rule is enabled. Disabled rules do not block any messages.

Hits

The number of spam messages that the rule blocked, a rough measure of how effective it is.

When SpamSieve checks whether a message matches a rule, it compares the part of the message named by the **Header** column with the contents of the rule's **Text to Match** column. The following are the message parts that may be used in the **Header** column:

From (address)

The e-mail address (not the name) of the message's sender.

To (any address)

The e-mail addresses of the primary recipients of the message. SpamSieve checks each recipient separately to see whether it matches the rule.

CC (any address)

The e-mail addresses of the carbon copy recipients of the message. SpamSieve checks each recipient separately to see whether it matches the rule.

Reply-To (address)

The address that you would be sending to if you replied to the message. This is often the same as the From address, but it could also be a mailing list or an alternate address for the sender. If the message does not specify a Reply-To, then the rule will not match.

List-ID

For mailing list messages, this hidden header indicates which mailing list the message was sent to.

List-Unsubscribe

For mailing list messages, this hidden header indicates how to unsubscribe from the mailing list. Some mailing list messages that do not have a List-ID header do have a List-Unsubscribe header.

Mailing-List

For mailing list messages, this hidden header indicates which mailing list the message was sent to. Some mailing list messages that do not have a List-ID header do have a Mailing-List header.

Received (any)

The Received headers contain information about the servers that relayed the message on its journey from the sender to the recipient.

Return-Path

This header contains information about where the message originated.

Subject

The subject of the message.

Body (any text part)

The content of the message. Some message contain more than one text part (for instance, plain text and HTML representations of the same message). The rule matches the message if any of the text parts matches the rule's text.

There are several different ways in which SpamSieve can compare the text in the message's header to the rule's text. In all cases, capitalization does not matter; lowercase letters are considered the same as their uppercase counterparts.

Is Equal to

The message matches the rule if its text is exactly the same as the rule's text. This is the fastest style of matching.

Contains

The message matches the rule if the message text contains the rule text.

Starts with

The message matches the rule if the message text begins with the rule text.

Ends with

The message matches the rule if the message text ends with the rule text. This is useful for matching domain names.

Matches Regex

This is like Contains, except that the rule text is treated as a Perl-compatible[21] regular expression[22]. Regular expressions are a powerful way of specifying patterns of text, for instance: e-mail addresses that contain numbers before the @ sign or subjects that are longer than 30 characters. If the regular expression entered in the **Text to Match** column is invalid, SpamSieve will color it in red, and it will not match any messages.

You can edit a rule's **Header** or **Match Style** by clicking in the corresponding column and selecting from the pop-up menu. To edit a rule's **Text to Match**, double-click the text. The SpamSieve Web site shows some examples of possible rules>[23].

[21] `http://pcre.org/pcre.txt`
[22] `http://zez.org/article/articleview/11/`
[23] `http://www.c-command.com/spamsieve/screenshots.shtml`

Training SpamSieve with a spam message adds its sender to the blocklist. You can delete a rule from the blocklist by selecting it and pressing Delete. You can copy the selected rules to the clipboard or drag and drop them into another application. You can type the first few letters a rule's **Text to Match** to quickly locate that rule.

5.2.7 Show Whitelist

The whitelist works the same way as the blocklist except that messages sent from addresses on the whitelist are *never* considered to be spam. The whitelist also has special support for mailing lists. If you train SpamSieve with a good message from a mailing list, it will add a rule to the whitelist that matches the message's mailing list header (List-ID, List-Unsubscribe, or Mailing-List). Then SpamSieve will know that all messages from that mailing list are good, regardless of who sent them.

The whitelist is most commonly used for matching messages sent from particular addresses, domains, or mailing lists. You can also use the whitelist to create *codewords*. For instance, you could create a rule in the whitelist that matches subjects containing "eggplant" (or some other word unlikely to occur in regular mail). You can tell select people to put "eggplant" in the subject of messages that they send you, and then you can be assured that their messages will get through, *even if the sender addresses are not in your address book or whitelist.*

5.2.8 Add Rule

This command creates a new rule in either the whitelist or blocklist. This is useful if you want to add your own rules to the whitelist or blocklist, rather than having SpamSieve learn the rules when you train it with messages.

5.2.9 Show Training Tip

This opens the **Training Tip** window, which gives advice for improving SpamSieve's accuracy, based on your current corpus and preferences.

6 Preferences

6.1 Filters

6.1.1 Order

SpamSieve uses a variety of filters to determine whether messages are spam or good. It consults the filters in the order listed in this window. When a filter decides that the message is good or spam, SpamSieve stops moving down the list. Thus, the order of the filters makes a difference. You can see from the order that if a message's sender is on the whitelist, it will be marked as good even if the Bayesian classifier would have predicted it to be spam. Normally this is what you want; the point of a whitelist is that you can be sure that certain messages will *never* be marked as spam.

6.1.2 Check for message in corpus

SpamSieve learns as you train it, but training is not instant. Training SpamSieve with a message will not necessarily give it enough information to classify that message correctly based only on the words in the message. However, once you have added a message to the corpus, SpamSieve *knows* whether it is good or spam, even though it might not make the correct prediction based on word probabilities. This option causes SpamSieve to see if it knows whether a message is good or spam before trying to calculate its spam probability. If SpamSieve has seen the message before, it will always classify it correctly. You can disable this option if you want to see what SpamSieve would have predicted if it did not already know whether the message was good or spam.

6.1.3 Use Mac OS X Address Book

With this option enabled, SpamSieve will never predict a message to be spam if its sender's e-mail address is in the system address book.

You can add addresses to the system address book using the Address Book application (located in the `/Applications` folder), or directly from an e-mail client that supports the system address book.

Mailsmith and PowerMail users should be sure to enable the option to use Apple's Address Book. Entourage users may prefer to use their address book as a whitelist instead of Apple's. This is described in the Entourage section.

6.1.4 Exclude my addresses

Enable this option so that spam messages with your own return address are not marked as good. Disable it if you send yourself messages and want to make sure that they are never marked as spam. SpamSieve looks on the "Me" card in Apple's Address Book to determine which addresses are yours.

6.1.5 Use Entourage address book

With this option enabled, SpamSieve will never predict a message to be spam if its sender's e-mail address is in the Entourage address book. When you start using SpamSieve, you should click the **Load** button to make SpamSieve read in the addresses in the Entourage address book. (This will cause SpamSieve to launch Entourage if it is not already open.) The addresses are loaded into SpamSieve's memory and stored in its preferences file, but they are not displayed in the **Whitelist** window.

Whenever you add addresses to the Entourage address book, you should go back to SpamSieve's preferences and click **Load** so that SpamSieve learns about the new addresses. Do not click **Load** while Entourage is downloading and filtering mail, as this may cause it to freeze.

6.1.6 Use SpamSieve whitelist

Enable this option so that messages sent from addresses on the SpamSieve whitelist are never marked as spam.

6.1.7 Use SpamSieve blocklist

Enable this option so that messages sent from addresses on the SpamSieve blocklist are always marked as spam.

6.1.8 Honor Habeas headers

The Habeas[24] service licenses a haiku to users who agree not to send spam e-mails. The users can then include the following text in their e-mails:

[24] `http://www.habeas.com`

```
X-Habeas-SWE-1: winter into spring
X-Habeas-SWE-2: brightly anticipated
X-Habeas-SWE-3: like Habeas SWE (tm)
X-Habeas-SWE-4: Copyright 2002 Habeas (tm)
X-Habeas-SWE-5: Sender Warranted Email (SWE) (tm). The sender of this
X-Habeas-SWE-6: email in exchange for a license for this Habeas
X-Habeas-SWE-7: warrant mark warrants that this is a Habeas Compliant
X-Habeas-SWE-8: Message (HCM) and not spam. Please report use of this
X-Habeas-SWE-9: mark in spam to <http://www.habeas.com/report/>.
```

When SpamSieve sees this text in a message and **Honor Habeas headers** is checked, SpamSieve assumes that the message is not spam. Habeas has promised to sue anyone who includes their haiku in a spam message, so many spammers are reluctant to do this. However, since some spammers have started including the Habeas headers, in order to get through spam filters, **Honor Habeas headers** is disabled in SpamSieve by default.

6.1.9 "ADV" messages are spam

This option causes SpamSieve to always mark messages as spam if they contain some variant of "ADV" at the start of the subject line. The "ADV" marker is used by some commercial bulk mailers.

6.1.10 Encoded HTML mail is spam

Many spammers encode the contents of their messages with base-64 so that filters cannot see the incriminating words they contain. SpamSieve can decode and look inside these messages. This option causes it to mark *all* such as spam, regardless of their contents, on the theory that legitimate senders do not try to obscure their messages.

6.1.11 Use Bayesian classifier

This enables SpamSieve main spam detector, which uses the corpus and word probabilities to identify spam messages.

6.2 Notification

6.2.1 What Is Notification?

All e-mail clients can notify you when you receive new messages, but most will notify you even when all the new messages are spam. If your e-mail client is not savvy in this way, you can turn off its notification and let SpamSieve notify you only when there are new good messages.

Note that if you have an Entourage rule set to only apply SpamSieve to messages that aren't from people in your address book, then SpamSieve won't count those messages as good for the purposes of notification (because it won't see the messages).

6.2.2 Play sound

This makes SpamSieve play a sound when new good messages are received. To add a sound to the menu, copy the sound file to the `Sounds` folder in your `Library` folder.

6.2.3 Bounce Dock icon

This makes SpamSieve bounce its Dock icon once when new good messages are received.

6.2.4 Keep bouncing

You might not be looking at the Dock icon when it first bounces, so this makes SpamSieve continue bouncing its Dock icon until you activate SpamSieve or your e-mail client.

6.2.5 Show number of new good messages in Dock

This option makes SpamSieve show the number of new good messages in its Dock icon. If there are no new good messages, SpamSieve will not show any number (rather than showing zero). Clicking the Dock icon, activating your e-mail client, or training SpamSieve with a message will reset the count.

The three sliders control the size and position of the number in the Dock icon.

6.3 Training

6.3.1 Allow duplicates in corpus

If you allow duplicate messages in the corpus, adding the same message twice will increase the counts for the words in that message. If you do not allow duplicate messages, the second and subsequent times you try to add a message will have no effect. By default, duplicate messages are not allowed in the corpus. This is nice because it means that you do not have to remember which messages you have already added; accidentally adding the same message more than once will not skew the data that you are providing to SpamSieve. If you wish to intentionally skew the data, you can check one or both boxes to allow duplicates.

6.3.2 Auto-train

These options cause SpamSieve to automatically train itself with newly received messages based on their predicted categories. Thus, after the initial training you would only need to train SpamSieve to correct its mistakes; it would automatically learn from all the other new messages.

Note that with either of these options enabled it is imperative that you correct SpamSieve when it makes a mistake; otherwise it will be making predictions based on incorrect information.

Auto-training with good messages tends to prevent false positives while slightly increasing false negatives. Auto-training with spam messages tends to prevent false negatives while slightly increasing false positives.

For best results, SpamSieve's corpus should contain several hundred messages, about 65% of them spam. Enabling auto-training for one or both types of messages can help build up the corpus to this desired state. SpamSieve works best when trained only on misclassified messages. Once it has attained a good accuracy, you should turn off auto-training all together.

6.3.3 Train SpamSieve whitelist

With this option enabled, training SpamSieve with a good message will add the message's sender to SpamSieve's whitelist. Training SpamSieve with a spam message will disable the sender if it appears in the whitelist.

Example: You receive an Amazon order receipt and train SpamSieve with it as a good message. This puts `auto-confirm@amazon.com` on the whitelist so that future order receipts are always marked as good. A spammer might decide that `auto-confirm@amazon.com` would make a good fake return address. If you receive such a spam, SpamSieve would mark it as good because the sender was on the whitelist. If you then tell SpamSieve that the message is spam, it will disable the whitelist rule so that it can catch future spam messages with that return address.

6.3.4 Train SpamSieve blocklist

With this option enabled, training SpamSieve with a spam message will add the message's sender to SpamSieve's blocklist. Training SpamSieve with a good message will disable the sender if it appears in the blocklist.

6.3.5 Train Bayesian classifier

With this option enabled, training SpamSieve with a message will add the words from that message to SpamSieve's corpus. It is highly recommended that you train the Bayesian classifier, as this is how most spam messages are caught.

6.3.6 Show training tip at startup

With this option enabled, SpamSieve will open the **Training Tip** window each time it is launched.

6.4 Advanced

6.4.1 Spam-catching Strategy

This slider lets you adjust SpamSieve's bias. The bias controls how aggressive SpamSieve is at catching spam. When SpamSieve is more aggressive, it is better at catching spam messages that share some characteristics with your good mail. When SpamSieve is more conservative, it will be better at marking borderline messages such as order confirmations and press releases as good. This setting is very powerful, and most users should stick to the middle range. It is also not a substitute for training SpamSieve. Only change the bias if SpamSieve is consistently making errors in the same direction.

6.4.2 Use full junk score range

With this option unchecked, Eudora will show a junk score of 0 for good message and 100 for spam messages. With this option checked, messages will be assigned scores *between* 0 and 100 depending on how spammy they are. Messages with scores of 50 or higher are considered to be spam. When you check or uncheck this checkbox, the change will not take effect until the next time you launch Eudora.

If you use the full junk score range, you must configure Eudora's **Junk Mail** settings so that its threshold is 50. The easiest way to do this is to click the **Set Junk Threshold** button.

6.4.3 Save false negatives to disk

False negatives are spam messages that SpamSieve didn't catch. This option causes SpamSieve to save such messages for later analysis. Clicking the **Show** button opens the folder containing the saved messages. You can e-mail this folder, or selected files from it, to spamsieve-fn@c-command.com[25]. By looking at the messages that SpamSieve missed, I can improve its algorithms to catch such messages in the future. Note that enabling this option will slow down SpamSieve's processing.

7 Frequently Asked Questions

7.1 Why is SpamSieve not very accurate for me?

SpamSieve is nearly 100% accurate, but only when properly trained. For best results, the corpus should have between 50% and 70% spam messages, as shown at the bottom of the **Statistics** window.

You can enable or disable auto-training in SpamSieve's preferences to help achieve the desired corpus ratio. For instance, if the corpus has mostly good messages, you can enable auto-training of spam messages to build up that side of the corpus. If the ratio and accuracy are already good, you can disable auto-training entirely.

The messages in the corpus should be representative of the messages that you receive. Adding more messages to the corpus generally improves accuracy, but it is not necessary to have more than a few thousand messages in the corpus.

[25] `mailto:spamsieve-fn@c-command.com`

7.2 How can I hide SpamSieve's Dock icon?

The easiest way is to use the free Dockless[26] utility. You'll need to make the Dock icon visible again in order to configure SpamSieve's preferences, view the statistics, or access any other part of SpamSieve's user interface.

As of this writing, Dockless is incompatible with Mac OS X 10.3. To hide the Dock icon the "hard" way, hold down the Control key and click on the SpamSieve icon in the Finder. Choose **Show Package Contents** from the menu. Open the `Contents` folder, and then open the `Info.plist` file. At the bottom of `Info.plist`, change:

```
<key>LSUIElement</key>
<string>0</string>
```

to:

```
<key>LSUIElement</key>
<string>1</string>
```

Then save the `Info.plist` file and relaunch SpamSieve. You may need to temporarily move the SpamSieve application to the desktop and double-click it there in order for Mac OS X to notice that you now want the Dock icon hidden. To make SpamSieve's Dock icon visible again, change the `Info.plist` file back; that is, change the `1` back to a `0`.

7.3 How does SpamSieve compare with Eudora's SpamWatch?

Although both use similar technology, SpamSieve's Bayesian classifier is more accurate and learns more quickly. SpamSieve is also more customizable. For instance, it lets you create a whitelist and a blocklist in addition to using the Bayesian classifier. If you look at the **Filters** tab of SpamSieve's Preferences window, none of the items above "Use Bayesian Classifier" are supported by Eudora. Lastly, SpamSieve can work with the Sponsored (free) Eudora, while SpamWatch requires the Paid ($50) version of Eudora.

7.4 Is SpamSieve 2.1.3 a free upgrade?

Yes, if you already registered SpamSieve you do not need to register it again. Your registration name and serial number will continue to work with 2.1.3.

[26] `http://homepage.mac.com/fahrenba/dockless/dockless.html`

7.5 Do you plan to support GyazMail, Mulberry, Netscape, NisusEmail, Thunderbird, or QuickMail?

At present these clients are not AppleScriptable enough to work with SpamSieve. GyazMail's developer is working on adding support for SpamSieve.

7.6 I'm using Eudora 6, but I don't see the Junk command in the Message menu. Where is it?

The Eudora installation instructions show how to enable this command from the esoteric settings.

7.7 How can I use SpamSieve with AOL?

If you already have a copy of Claris Emailer, you can configure it to access your AOL account and filter the mail using SpamSieve. If you do not already have Emailer, it may be hard to find because it was discontinued a few years ago. Instead, you can use a utility such as Mail Forward[27] to forward your AOL mail to another account. Then you can download the messages using one of the e-mail clients that SpamSieve supports directly.

7.8 What information should I include when I report a problem?

If you are reporting a problem with SpamSieve's accuracy, open the **Statistics** window and choose **Copy Stats** in the **Edit** menu so that you can paste all your statistics into the message. Also, use the **Open Log** command in the **Filter** menu and include any relevant entries from the log.

7.9 Why does SpamSieve try to connect to dreamhost.com when it starts up?

It's checking to see whether there's a newer version of SpamSieve available. You can disable this feature from the **Software Update** window that's accessible from the **SpamSieve** menu.

[27] `http://www.sspi-software.com`

7.10 Where can I download the older Mac OS 9 version?

There has never been an OS 9 version of SpamSieve—sorry.

7.11 Can I delete spam messages after training SpamSieve with them?

Yes. When you train SpamSieve, it stores the information from the messages in its own data files. It is not necessary to keep the spam messages in your e-mail client.

8 Purchasing and Support

8.1 Contact Information

You can download the latest version of SpamSieve from the SpamSieve Web site[28]. Questions about SpamSieve may be sent to spamsieve@c-command.com[29]. 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 News mailing list[30]. The traffic on this list is very low, only one message per new version of SpamSieve.

8.2 Purchasing

SpamSieve has a free trial period that lasts for 30 days or 20 launches, whichever is longer. To use SpamSieve beyond the demo period, you must purchase it. This entitles you to free updates and support.

To purchase, choose **Purchase**... from the **SpamSieve** menu. You can use the **Instant Purchase**... button to enter your billing information from within SpamSieve or use the **Web Purchase**... button to enter it from your Web browser.

[28] `http://www.c-command.com/spamsieve/`
[29] `mailto:spamsieve@c-command.com`
[30] `http://www.c-command.com/spamsieve/support.shtml`

Soon after paying, you'll receive an e-mail with your serial number. If you used **Instant Purchase**. . ., you're done. If you used **Web Purchase**. . ., enter the name and serial number from the e-mail into the **Purchase** window and click **Personalize**. If you need to re-install SpamSieve, you can simply re-enter your name and serial number and click **Personalize**; there's no need to purchase again.

If you purchased SpamSieve but cannot find your serial number, click the **Lost Your Serial Number?** button. This will open a form where you can enter your e-mail address and look up your order information.

A license for SpamSieve is good for one person *or* one computer. You can install it on one Mac, and everyone sharing that Mac can use it (on that Mac). Alternatively, you can install it on your desktop Mac and your PowerBook; you can then use it on either machine, provided that no one is using it on the other machine.

SpamSieve uses eSellerate Product Activation[31] to reduce software piracy. This should be completely transparent except that you will need to be connected to the Internet when you first enter SpamSieve's serial number. (Subsequent launches do not require an Internet connection.) eSellerate's privacy policy is as follows:

> eSellerate Product Activation is an anti-piracy technology that publishers can use to protect the software they sell through eSellerate.
>
> During activation, eSellerate looks at the computer's present configuration and uses that data to create a unique hardware identification. The unique hardware identification does not include any personal information, nor does it include any information about the software or documents that reside on the computer. The hardware identification identifies only the computer's configuration, and is used only for activation purposes.
>
> Once software is activated on a computer, minor changes to that computer's configuration will not affect the activation. Major changes to the computer's configuration may require reactivation of the software. Reactivation is at the publisher's discretion.

Purchasing SpamSieve entitles you to a reasonable number of activations. You can activate SpamSieve on your desktop, on your laptop, and on new Macs that you buy in the coming years. If you run out of activations, e-mail me and I'll most likely give you more. The goal is to make things as easy as possible for owners of SpamSieve.

[31] `http://www.esellerate.net/papolicy.asp`

8.3 Legal Stuff

[32] `mailto:mjt@c-command.com`
[33] `http://www.pcre.org`
[34] `http://www.mulle-kybernetik.com/software/EDFrameworks/download.html#EDCommon`
[35] `http://www.mulle-kybernetik.com/software/EDFrameworks/download.html#EDMessage`

9 Version History

2.1.3—April 7, 2004

- Added menu commands for installing the Apple Mail plug-in and scripts, the Eudora plug-in, and the Entourage scripts. These items are now stored inside the SpamSieve application package.
- Scripts for the other applications are now stored inside the application bundle, not at the root of the disk image. The **Show Other Scripts** menu command will reveal them in the Finder.
- Apple Mail and Eudora users should update their plug-ins, using the commands in the **SpamSieve** menu.
- Added the `score` script command, which returns an integer between 0 and 100 indicating how spammy the message is. 50 and higher mean spam.
- The `Predicted` lines in SpamSieve's log now show the scores of the messages.
- Can now use Eudora's full 0-to-100 junk score range if you check the appropriate box in the **Advanced** preferences.
- Improved parsing of messages with 8-bit transfer data.
- Faster at processing messages.
- Added support for Outlook Express 5.
- Worked around OS bug that could cause SpamSieve to come to the front each time a message was processed in Apple Mail or Eudora (usually if an X11 application was frontmost).
- Fixed bug where errors encountered while processing messages were not reported in the log.
- Worked around Cocoa problem where certain notification sounds wouldn't play.
- Made the **Purchase** window easier to understand, and added a button for looking up lost serial numbers.
- Trims the text in the serial number field so people don't accidentally paste the number in twice.
- Software updater is better at checking whether the computer can connect to the Internet.
- SpamSieve now tries to parse Eudora messages according to RFC822, even though this will sometimes fail, as many Eudora messages are not RFC822-compliant.
- Adjusted the list of headers that SpamSieve ignores.
- Added keyboard shortcuts for Apple Mail scripts.
- Improved the training tips.

- Updated to SQLite 2.8.13.
- Updated to eSellerate SDK 3.5.5.
- The **Send Report** button in the crash reporter is no longer a default button, so there's no longer confusion about entering returns in the comment field.
- No longer prints fragments of spam messages to the console when it gets confused.
- Replaced the copy of the manual outside the app with a readme.

2.1.2—January 26, 2004

- SpamSieve can now move Apple Mail POP messages to the **Spam** folder. Thus, it now fully supports Apple Mail on Jaguar and Panther.
- **Honor Habeas headers** is now off by default.
- Fixed regression where blocklist and whitelist rules got deselected after editing their text.
- When loading addresses from Entourage, SpamSieve now picks up addresses that are not associated with any contact (that is, they appear only in a group).
- The default date shown in the **Statistics** window is now the date that SpamSieve was first launched, rather than September 2002.
- The Apple Mail `Add Good` script is better at finding the proper inbox when moving false positives out of the **Spam** folder.
- The Apple Mail `Mark If Spam` script can mark the spam messages as read.
- The **Purchase** window now makes it more clear when a serial number has been accepted.
- In the **Statistics** window, **Set**... is now **Set Date**... and **Copy** is now **Copy Stats**.
- Fixed crash that could happen when processing messages in Japanese encodings.
- Added Japanese localization.

2.1.1—January 8, 2004

- Much faster at processing messages when there are many blocklist and whitelist rules. Also improved the speed of loading, deleting, and sorting rules.
- Improved accuracy tracking with the Panther version of Apple Mail; previously, SpamSieve couldn't always tell when it was being corrected.
- Catches more spam because it knows about more spammer obfuscation tricks and also which headers it should ignore.
- Fixed bug (introduced in 2.0) where the Bayesian engine didn't work if Mac OS X's default language was set to Japanese.

- The SpamSieve Eudora Plug-In is better at launching the SpamSieve application if it is not already running.
- Loading Entourage addresses now adds to the addresses that were previously loaded, rather than replacing them. This makes it possible for Entourage users who have more than one Entourage identity to give SpamSieve the addresses from all their address books (by loading once for each identity). Hold down Option when clicking Load to get the old behavior of replacing the previously loaded addresses.
- The sound pop-up menu in the **Preferences** window now immediately notices when new sounds are installed; previously, it would only check when updating the rest of the preferences window.
- You can now add a rule without a the **Blocklist** or **Whitelist** window being frontmost. SpamSieve will ask which type of rule to add.
- Regex rules can now start with an options modifier such as `(?-i)`.
- Copying rules to the clipboard now just copies the text to match (typically an e-mail address), not all the columns. To get all the columns, you can print to PDF.
- The Entourage `Add Good` script now finds localized inboxes, rather than creating a folder called **Inbox**.
- The Entourage `Add Spam` script can now remove spam messages from the server.
- **The Statistics** window now shows percentages instead of ratios.
- SpamSieve will now quit at launch if another copy of the application is already running.
- Re-targetted broken Habeas URL.
- Added the following menu commands: **Close All Windows**, **Minimize All Windows**, and **Zoom**.

2.1—December 9, 2003

- Added support for Apple Mail POP accounts. POP messages can be marked as junk and colored, but (due to limitations in the present version of Apple Mail) they cannot be moved to another mailbox.
- Added a **Training Tip** window that gives advice on how to improve SpamSieve's accuracy, based on the current state of the corpus and preferences.
- Rules in the whitelist and blocklist are no longer limited to just matching sender addresses. They can now match a variety of message fields (To, CC, Subject, etc.), as well as the message body. In addition to exact matches, rules now support the following match styles: contains, starts with, ends with (useful for matching domains), and Perl-compatible regular expressions. You can now edit

rules and add new rules manually (as opposed to automatically, as a result of training SpamSieve with a message).

- When trained with a good message from a mailing list, SpamSieve will automatically create a whitelist rule based on a mailing list header, if present.
- SpamSieve can now read in the Entourage address book and use it as a whitelist. Thus, the Entourage rule can now give SpamSieve all the messages, not just the ones that were from unknown senders. This means that SpamSieve can now accurately notify the user when non-spam messages are received. Also, the statistics it keeps will be more complete.
- Improved the accuracy of the Bayesian classifier when the corpus is unbalanced.
- Made a variety of low-level changes to improve SpamSieve's accuracy, for instance: adjusted the list of headers that are analyzed and how words are tokenized.
- The Apple Mail `Add Spam` script now has an option to control whether the messages are moved to the **Spam** folder.
- The Entourage `Add Good` script now moves messages to the inbox if they're located in the **Spam** folder.
- Mailsmith users can now auto-train using only spam or good messages by turning off training in Mailsmith and turning on one of the auto-train checkboxes in SpamSieve.
- Improved the importing of mbox files that do not have blank lines between the messages, such as some Eudora mailboxes. Fixed a bug where the the mbox parser could crash if a message had length zero. Also, SpamSieve now shows a progress bar while counting the number of messages that will be imported.
- Improved the corpus and rule list displays. You can now enter and leave editing mode by typing Return. Typeahead works better; for instance, if you type "g" and there are no rows that start with "g," it will look for one that starts with "f." When you delete a word or rule, you can cancel out of the confirmation sheet by typing Escape. To avoid the confirmation sheet entirely, you can delete using Command-Delete instead of Delete. When a word or rule is deleted, SpamSieve selects a nearby rule so that you don't lose your place. When deleting many words at once, SpamSieve no longer shows a progress window for deletions that will not take very long.
- Entering the name and serial number to personalize SpamSieve is now more foolproof: SpamSieve strips leading and trailing whitespace, and it detects when you enter a coupon code in the serial number field. Fixed regression where SpamSieve rejected names containing non-ASCII characters. In addition, there's a new button for quickly redeeming coupons.
- Updated to the latest eSellerate SDK so that purchasing SpamSieve from within the application is faster.

- SpamSieve now requires Mac OS X 10.2.6 or later.
- Fixed bug where dates entered in the **Statistics** window were sometimes parsed in GMT instead of the local time zone, thus causing the date to be off by a few hours.
- Improved the reliability and user interface of the crash reporter.
- No longer crashes when parsing certain non-RFC822-compliant Eudora messages.

2.0.2—October 1, 2003

- Now works with Apple Mail (IMAP and .Mac only, not POP).
- The message count in the Dock icon now resets when an e-mail client becomes active (rather than just when SpamSieve became active). You can also control the size and position of the number in the Dock icon.
- Much faster at deleting lots of rules at once.
- Replaced the message store database with custom code that's faster and more reliable.
- Improved accuracy for HTML messages containing links.
- Importing mbox files is faster.
- Fixed bug where you couldn't use Web registration after the demo period had expired.
- The log records which addresses matched the whitelist or blocklist.
- The log records corpus imports.
- Auto-training is faster.
- The Entourage Add Spam script can close the frontmost window if it's spam.
- Fixed bug where the date in the **Statistics** window could get cut off if you changed it to use a more verbose format.
- Worked around OS bug that caused dates like "01.09.2003" to be interpreted as January 9 in German-style locales.
- Added **Copy** button to the **Statistics** window.
- Fixed problem updating certain history databases from 1.3.1.
- More resilient to minor corpus file corruption.
- Fixed crash that could happen with improperly formed multi-part messages.
- Shows the number of blocklist or whitelist rules in the title bar.
- The whitelist now contains some c-command.com addresses by default.
- Fixed crash when opening the **Statistics** window while adding messages.
- The **Statistics** window shows ratios, where applicable.

- Assorted minor performance improvements.
- The modification dates of the AppleScripts are now the actual modification dates, not the date the distribution was built.

2.0.1—September 17, 2003

- Replaced the database engine that was being used to store the corpus with some custom code. This should be much faster and more reliable.
- Loading and saving the rules is faster, due to a better file format.
- The rules and corpus message counts are now saved to disk during idle time rather than when quitting. This should prevent data loss in the event that SpamSieve doesn't quit normally.
- The whitelist and blocklist are more memory-efficient.
- Plugged memory leak in EDMessage.
- Fixed crash involving certain really long header lines.
- Fixed bug where the **Whitelist** and **Blocklist** windows weren't always up to date.
- Table views are smarter about not scrolling unnecessarily to maintain their selections.
- The **Whitelist** and **Blocklist** windows now secondary sort by sender.
- Changes to the preferences are saved to disk immediately.
- Fixed bug where tables saved their sorted columns but didn't restore them.
- When SpamSieve gets a fatal error, it now quits like it says it will.

2.0—September 10, 2003

- SpamSieve now extracts *a lot* more information from each message. This makes it much more accurate and also makes it learn faster.
- Now integrates with Eudora 6 (Sponsored or Paid) via a plug-in. It can now process every incoming Eudora message and can be trained using the **Junk** and **Not Junk** commands in Eudora's **Message** menu.
- SpamSieve now has a blocklist and a whitelist. These are automatically maintained based on the senders of messages that SpamSieve is trained with. The blocklist makes sure that all messages from known spammers are caught and speeds processing for these messages. The whitelist lets you be sure that certain messages will never be marked as spam; this was possible before, but now you don't have to clutter your address book with addresses from online retailers, etc.
- You can now control how conservative or aggressive SpamSieve is at catching spam.

- SpamSieve can now play a sound or bounce its Dock icon after a batch of non-spam messages has arrived. This is meant to replace your e-mail client's new mail notification, which you don't want going off if all the new messages are spam.
- Shows the number of new good messages in the Dock icon.
- Now parses HTML so that it can better extract relevant information from HTML messages, and also handle various HTML-based tricks that spammers use to fool filters.
- New method of calculating word probabilities makes SpamSieve better at discerning which words in the message are important.
- Includes a corpus of seed spam, to jump-start spam recognition for users who do not have many saved spam messages.
- The corpus is now stored in databases rather than in a property list. This makes it launch faster and use much less memory, as the corpus doesn't have to be all in RAM at the same time.
- The statistics file format (for History.db) has changed in order to enable performance improvements and more statistical displays in future versions.
- Handles more types of plain text obfuscations, and is much faster at undoing them.
- Added option for the address book whitelist to only use other people's addresses, so that spam messages from your own address don't match the whitelist.
- Can mark all messages with Habeas headers as good.
- Can mark all messages with some variant of "ADV" at the start of the subject as spam.
- Can mark all base64-encoded HTML messages as spam.
- New probability combiner increases accuracy.
- Uses stop words to speed processing and reduce false negatives.
- When filtering a message, considers the number of occurrences of the words, not just which words are present.
- Can import messages from mbox files.
- Can import the corpus from and export it to an XML property list (the same format used by 1.x).
- SpamSieve can now check for updated versions of itself.
- Added crash reporter.
- Added Dock menu containing frequently used commands.
- The entries in the log are more detailed.
- The corpus now stores the date at which each word was last accessed.

- Fixed bug where storing statistics would fail on systems that didn't know about GMT.
- Fixed bug where SpamSieve could throw away long runs of HTML thinking they were attachments.
- Added button for opening the Mac OS X Address Book from inside SpamSieve.
- The **Statistics** window now has a contextual menu item for copying the displayed information.
- SpamSieve no longer wastes cycles updating the **Statistics** window after it's been closed.
- The **Statistics** window is smarter about updating only the portions that could have changed.
- No longer shows Good Words and Spam Words stats.
- Logging has less overhead.
- Updates the history asynchronously, resulting in faster message processing.
- Checks for mistakes in a background thread.
- False negatives are now written to disk in a background thread.
- Re-arranged the **Corpus** window.
- Pruning the corpus now works by access date rather than by word counts. Of course, you can manually prune the old way by sorting the **Corpus** window by **Total**.
- Updated to SQLite 2.8.6 and tuned it for speed.
- Updated to PCRE 4.3.
- Updated to eSellerate 3.5, which should fix crashes some people saw after registering on 10.2.6.
- Now looks at headers of subparts of messages from Mailsmith.
- Time-consuming operations now either have a progress bar or a progress spinner.
- Better at extracting malformed e-mail addresses from headers.
- Copying rows from the **Corpus** window to the clipboard now uses the order of the columns in the window rather than the default column order.
- Fixed regression where the Entourage scripts no longer created the **Spam** folder if it didn't exist.
- Fixed potential crash with regex replacements at the end of a string.
- History.db and the corpus can now be aliases.
- Automatically trims carriage returns and other illegal characters when you paste in your name and serial number.

- Now saves the name and serial number to disk as soon as they're entered.
- The **Spam** folder in Entourage no longer has to be top-level.
- Entourage can mark good messages as unread.
- Type-selecting in table views is quicker.
- No longer nags constantly when unregistered.
- Fixed bug where it could *look* as though SpamSieve had hung if it started up in the background with an empty corpus.

1.3.1—June 18, 2003

- Added direct integration with Mailsmith 2.0 and later. Enabling SpamSieve is as easy as clicking a checkbox. You can train SpamSieve directly from Mailsmith's Message menu. Bare Bones Software has seamlessly integrated it with Mailsmith's powerful filtering system, and Mailsmith knows not to bounce its Dock icon after receiving a batch of messages that are all spam.
- Fixed crashing bug triggered by incorrectly encoded headers.
- Regex substitutions are faster and much more memory efficient.
- When adding spam messages to the corpus, the default is now for SpamSieve to move them to the **Spam** folder.
- The PowerMail **Move If Spam** script now changes the color of spam messages.
- The Emailer scripts now pass text and HTML attachments on to SpamSieve for analysis.
- Added instructions for using the Entourage and PowerMail address books as whitelists.
- Compacted the ED frameworks to reduce application size and memory use.
- Disabled SQLite's file locking so that SpamSieve's data folder can now be located on an AppleShare volume.
- Caches the Address Book to speed whitelist lookups 100 fold.
- The statistics database is faster due to an updated version of SQLite.
- Fixed bad pointer in header extractor that caused unnecessary fallbacks to the secondary parser (reducing accuracy).
- Added experimental support for moving Entourage messages on IMAP accounts to the (local) **Spam** folder. This can be enabled by editing the scripts with Script Editor.
- For clarity, the names of AppleScripts that ask SpamSieve to predict the category of a message now contain the word "if."
- Added lots of minor clarifications to the documentation.

1.3—February 11, 2003

- More resilient to spammers' tricks for obfuscating words.
- Can use e-mail addresses in the system Address Book as a whitelist. Messages sent from those addresses will never be marked as spam.
- Greatly reduced overall memory usage as well as launch and quit times.
- Can save false negatives to disk for later reporting to SpamSieve's developer.
- You can edit the spam and good counts associated with a word, remove selected words from the corpus, and reset the corpus entirely.
- Type-ahead navigation in the **Corpus** window. Type the first few letters of a word or number to select it (and scroll to it).
- You can hide statistics from before a set date, to better see the current accuracy and spam reception rate.
- Improvements to the **Corpus** window: Shows all words rather than only those considered statistically significant. Re-sorting by numeric columns is twice as fast. You can copy the selected rows to the clipboard or drag them to another application. The selection is preserved when you change the sort column, you can sort in descending order, and the sorted column is remembered between launches. The Home and End keys work.
- The **Prune Corpus** command now tells you how many words it would remove and asks for confirmation.
- The statistics tracking is smarter about handling duplicate messages.
- The statistics have tooltips explaining what they mean, and you can copy all of the statistics to the clipboard at once.
- Improved accuracy tracking of PowerMail and Emailer messages.
- Eudora Integration: Can mark spam messages as read and/or mark them for removal from the server.
- Expanded the AppleScript dictionary, to enable better integration with mail and news clients.
- Entourage Integration: Creates Junk category if there isn't one, and can mark spam messages as read.
- Mailsmith Integration: The adding scripts now set the appropriate message properties.
- Better parsing of messages with illegal characters in the headers.
- SpamSieve's Info.plist file contains an `LSUIElement` entry. Change the `0` to a `1` to hide the application's Dock icon. (You'll need to change it back to access the preferences.)

- The message count display has moved from the **Corpus** window to the **Statistics** window.
- Better error message when the corpus couldn't be saved.
- Added tooltips to preferences.
- The registration window gives better feedback when you personalize.
- Better recovery from errors in the corpus file.
- The secondary parser is better at handling DOS linebreaks.

1.2.2—November 20, 2002

- Fixed bug in the PowerMail **Add Good** script.
- Added uninstaller for Eudora users.
- Better handling of errors while adding messages to the corpus.
- Removed bloat from the Entourage **Mark Spam** script.
- The application icon now has an alpha channel, so it doesn't appear with a white halo when viewed on a colored background.
- Minor changes to the manual.

1.2.1—November 18, 2002

- Modified Info.plist to work around a bug in Mac OS X 10.1 that could cause the Finder to crash when launching SpamSieve.

1.2—November 18, 2002

- Added support for Emailer 2.0v3 and Eudora (5.2 and later).
- Decodes base64 and quoted-printable text parts, thus finding words that spammers try to hide from anti-spam software.
- Decodes subjects that use different character sets (e.g. big5).
- Adds special tokens for MIME entities such as part boundaries and uninterpretable message parts.
- Keeps track of the messages added to the corpus, and can optionally prevent you from adding the same message more than once (biasing the counts). Thus, you no longer have to remember which messages you've already added.
- You can now "undo" adds to the corpus, e.g. if you added a message as good when you meant to add it as spam.
- Can now add messages to the corpus as they are filtered, so after the initial training you only have to add messages when SpamSieve makes a mistake.

- When filtering a message, SpamSieve can optionally check whether the message is in the corpus. If it is, SpamSieve looks up the answer rather than trying to predict. One use of this feature is that if SpamSieve makes a mistake, you can **Add Spam** and then **Label/Move If Spam** and be sure that the message will be labelled/moved.
- Keeps a log of additions to the corpus, filtering results, and errors.
- Mailsmith: If SpamSieve thinks a message is spam, it sets the **deleted** property of the message to true; otherwise it sets the **flagged** property of the message to true. Therefore, if SpamSieve has classified the message then exactly one of the properties will be true, and if it hasn't they'll both be false. (Normally, neither of these message properties is used by Mailsmith itself.)
- Entourage and PowerMail: If you tell SpamSieve to move spam messages to a spam folder and the spam folder doesn't exist, the script will create the spam folder for you.
- When you add spam messages to the corpus, can optionally move them to a **Spam** folder.
- Added status indicators in the Dock icon (like Norton DiskLight).
- The spam probability of unknown words is now 0.4 instead of 0.2.
- The **Corpus** window uses less memory and sorts much faster.
- Accuracy tracking is faster and uses less memory and disk space.
- Fixed bug where accuracy tracking didn't work for some Mailsmith messages with multiple parts.
- Improved the manual's instructions for e-mail client integration.
- Compiled with GCC 3 for greater speed.
- Uses the latest version of the eSellerate SDK, which eliminates a crash at startup under certain circumstances.
- No longer shows the "Upgrading From 1.0" message when starting with a blank corpus.

1.1—September 19, 2002

- E-Mail Client Integration
  - Added support for PowerMail.
  - Added instructions and an AppleScript for making Mailsmith download and filter mail faster.
  - Added an AppleScript for Entourage that moves spam into a Junk folder.
- Performance
  - Launches about 60% faster than 1.0.

  - You can now prune the corpus to remove words that are taking up memory without contributing to spam recognition. This can also dramatically decrease SpamSieve's launch time.
  - Recalculating spam probabilities is about 10% faster and uses less memory.
  - Quitting is faster because SpamSieve now writes corpus changes to disk during idle time.
  - Saving the corpus is slightly faster.
- Displays statistics about the number of messages filtered, SpamSieve's accuracy, and the types of words in the corpus.
- SpamAssassin's X-Spam-Status headers are now treated as single words. This means that if SpamAssassin is running on your mail server, SpamSieve will learn to respect (or ignore) its judgement.
- Does a better job of ignoring e-mail attachments, thus reducing corpus bloat.
- Installs the eSellerate Engine if it's not present, thus enabling "Instant Registration" for more users.
- Asking SpamSieve to categorize a message now forces an update of all the word probabilities. Previously, the update only happened during idle time.
- Highlights the sorted column in the **Corpus** window. The columns themselves have shorter names. There's a new "Total" column. Auto-resizing of the columns works better. You can now manually resize any column, and manual resizings and reorderings are saved between launches.
- Shows fatal errors as alert panels rather than just printing them on the console.
- The Corpus.plist data file is now sorted by word. This makes it easier to examine the corpus manually, and to compare it to other users' corpuses.

1.0—September 10, 2002

- First public release.