

ClaVettingProcess

Vetting Process for SpamAssassin CLAs

We started by identifying all the names of people who had contributed code, rules or documentation to SpamAssassin since the project was founded.

This was to identify the full list of contributors to the open-source code base, including patches that were trivial, affected a very small number of lines of code, or did not add functionality (e.g. a bug fix etc.).

We assembled this from several sources:

- the CVS logs of the code
- comments in the code itself
- data in the bug tracking system
- mailing list traffic from the Spamassassin-talk and Spamassassin-devel lists
- <http://spamassassin.org/credits.html> - the "credits" web page for the project

Note for NAI readers: this differs from the NAI process, in that we assembled a list of *all* contributors from the entire project lifetime. In other words – it was not performed based on what was in the lines of code we wanted, since we needed to "cleanse" the entire codebase.

Initial list was: <http://taint.org/wk/SAContribList?action=recall&date=1065743884>

The list was maintained at <http://taint.org/wk/SAContribList> – revision history on that page is here: <http://taint.org/wk/SAContribList?action=info>

We then went through the list, searching for email addresses for these users, and narrowed it down a little as it became evident that some of the "credits.html" users had not contributed any code. – Craig had added their names from the earliest list of names of people who had opened bugs in Bugzilla, believing that they'd contributed code. In some cases, they had not contributed code; just a bug report. He said:

These are probably mostly names I added to the credits page while at Deersoft -- I'll go through hopefully this evening and try and pull together how they all made the list. The criteria were pretty loose, so many of these will probably migrate to the removed section at that point - these are probably mostly people who filed bugzilla tickets as of some date back when.

In this file, the users are listed one-per-line, with a symbol indicating their status, their name and email addr, and a piece of text describing what they contributed or comments regarding this contribution.

By 2003-11-03, we had narrowed it down sufficiently, and needed to make it more easily shareable – so we moved it to a private CVS repository here: <http://cvs.sourceforge.net/viewcvs.py/spamassassin/asfpurgatory/spamassassin.csv> The current, most up-to-date list is at the "view (text)" link on that page.

The format of this file is a comma-separated values list, with each line like so:

```
code_removed,Ryan Cleary,<tryanc(at)interdimensions.com>,no,DCC support 2-line bug fix. no longer in codebase.
2002-07-08.; not in the distro anymore
```

The first word is the status of the contribution: "code_removed" for code that was removed from CVS so no CLA required, "third_with_cla" for third-party contributors with a CLA on file, "reimplement" for code that was removed and needs to be reimplemented from scratch, "third_no_cla" for third-party contributions without a CLA yet, "none_needed" for cases where there was no contribution that required a CLA, etc.

Next comes the name and email address, followed by a "where is the CLA" field used to track CLAs in transit, and finally, a textual description of what the change was, where it went, when, and whether it is still required.