# All.Net Analyst Report and Newsletter

***Welcome to our Analyst Report and Newsletter***

## Spamalot

Spam filters have gotten better and better at identifying spam (I don't know how to define it, but I know it when I see it?), and worse and worse at not identifying legitimate messages as spam. That is, lower false negatives, and higher false positives.

## I like spam filtering

I have grown my own, evaluated industry approaches, and done so since the mid-1990s. I like the fact that I don't have to look at it all, and it reduces denial of cognitive services (my brain) in favor of automation. It catches a lot of stuff I don't want to see unless I feel like looking for it. But, like anything else, too much of a good thing...

## But it's gone too far

Modern spam filtering seems to have forgotten or ignored what was used for many years and successfully. They should put back in the rule-based systems as a starting point and only go to the other methods afterwards. Here are some rules they should use as a first pass filter:[1]

- If you send something to me, it is not spam when I send something back to you.
  - Anything address I send to should be allowed to send to me, and if I specify an address as not allowed to send to me, I should not be allowed to send to it either.
- If you say emails from me are not spam, they are not!
  - I have had examples of people who have explicitly identified emails from me as not being spam and whose spam filters have sent them to spam even then. An allow list should be allowed and followed!
- Even if 500 people say something or a sender is spam or a spammer, it doesn't mean they are sending spam to everyone else.
  - Spam listing services should be treated as only one parameter in a complex decision, and never be treated as authoritative sources. They tend to take a small number of reports where someone hit 'spam' because they didn't want the email, instead of, for example, checking if there is a list removal URL and using it...
- No matter how many copies of the same content are sent over a short time period to different recipients, that does not make it spam.
  - I have mailing lists with 30,000 email addresses and they are all people who have signed up or agreed in one way or another to get the emails. None of these should ever be indicated as spam, and of course, if they don't want the emails, they can remove themselves.
- Just because a message is short, that does not make it spam.
  - I very often answer with 'Yes' or 'No'. Length is not an indicator of legitimacy.

1 For the purposes of this article, when I say something is or isn't, should or should not, I am not stating "it is my opinion that…" in front of each one. These are only my opinions.

  

- Even if messages use words or sequences you don't like, that doesn't make it spam.
  - If I use words on your bad list, that doesn't make it spam. In fact, I have had many instances where I quoted examples of spam in my messages, and those messages are not spam either.
- Individual preferences count, but don't guess you know what I want based on what I did or did not call spam before.
  - As I press the spam button on more and more on a domain that sends email from multiple addresses to multiple addresses, the system should learn that most common factors and use them as differentiators. But figuring out which of them are the real differentiators requires that you ask me instead of guessing with statistics.
- Spam should not be answered with spam.
  - When I click spam on emails, the provider (Google in my case) should not send me an email telling me that one of my users has clicked the spam button. That's just adding more crap to the crap. Be careful of positive feedback loops!
- Explain why you declared something spam and let me correct you.
  - Everything identified as spam should have the reasoning behind it made transparent to the recipient. And the recipient should be able to adjust the legitimate reasons to identify it as spam or not, so the user gets to choose – any time they want, and not as forced by the provider.
- One person's garbage is another person's gold.
  - Different strokes for different folks. Spam is not one size fits all. In fact, it's not even one size fits me. It is the recipients AND senders you are serving, not yourself, and not only the recipients. YOU don't know what's best for ME.

**Conclusions**

The spam filtering world has gone too far away from its roots in that:

- It no longer pays attention to obvious indicators of legitimate email, producing excessive false positives.
- It ignores obvious indicators of spam, like that emails from the same sending domain to the same recipient (or groups of them) previously marked as spam repeatedly.
- It ignores user preferences or makes them less than simple to express.
- It fails to explain WHY it declares things as spam and allow the user to determine whether that is right for them.
- It makes it inherently harder to get OFF a spam list than to get ON it, leading to abuse by folks who 'get even' for perceived dislikes by using spam lists as cudgels.
- Crowds are not wise, they are often stupid. With modern technology you should individualize spam filtering and make it very easy for the individual to control.

Fix these before rushing to further methods that are less reliable, and then only use the less reliable methods when the more reliable ones leave the answer unclear.

  