

microsoft.public.outlook.general: Re: Outlook should process rules first, before junk.

Re: Outlook should process rules first, before junk.

Source: <http://www.tech-archive.net/Archive/Outlook/microsoft.public.outlook.general/2004-09/5442.html>

From: Vanguardx (*see_signature*)

Date: 09/21/04

Date: Tue, 21 Sep 2004 16:00:45 -0500

"corpcon" <corpcon@discussions.microsoft.com>
wrote in news:19E0CB91-CE8E-492A-BCF3-A787EDAA8C69@microsoft.com:
> *You may be thinking of previous versions of Outlook that had the*
> *"Junk" and "Adult" filters. I know of no such thing in 2003.*

The OP never identified WHICH version of Outlook they were using, so any response regarding any version is a valid response. When you ask for a tire at the car shop and don't identify the specific car then any tire they give you satisfies your request, even a tire for a wheelbarrow.

I realize Microsoft wants to hide that they employ Bayesian filtering in OL2003 but there is little you can configure for that filtering. You cannot teach the filter before employing it by feeding it sets of good and bad e-mails (i.e., there is no learn or train mode) to reflect YOUR e-mail history, you cannot configure word expiration (to reduce the noise floor for words that fade out of your vocabulary or are not used for extended periods), it may be more susceptible to bias by using a flood of innocuous word lists in messages (I'm still looking for proof that good-word flood lists will negatively bias Bayesian filters *if* properly configured regarding expiry), and, in fact, there is almost nothing configurable regarding the Bayesian, er, Junk filter used in OL2003.

>From <http://office.microsoft.com/en-us/assistance/HP052429671033.aspx>:

"There are two parts to the Junk E-mail Filter: the Junk E-mail Filter Lists and the state-of-the-art technology developed by Microsoft Research that evaluates whether an unread message should be treated as a junk e-mail message based on several factors, such as the time it was sent and the content of the message. The filter does not single out any particular sender or type of e-mail message. The filter is based on the content of the message in general and uses advanced analysis of the message structure to determine the probability that it is a junk e-mail message."

Re: Outlook should process rules first, before junk.

microsoft.public.outlook.general: Re: Outlook should process rules first, before junk.

All that fluff just so Microsoft doesn't divulge that they use their own variation of Bayesian filtering. Geez. I wonder why Microsoft slapped the article at <http://www.mapilab.com/news/0042.html> (so the linked article isn't available anymore). Yeah, we wouldn't want to divulge how Microsoft did it although there are other open source Bayes filters available. Unfortunately I never got to see the article before Microsoft forced its deletion, but articles like <http://support.gfi.com/manuals/en/me10/me10manual-1-12.html> indicate that OL2003's spam filtering is Bayesian and comes with a preset level of training (so you cannot adapt the filter's statistical weighting towards YOUR sampling of good and bad messages –until– you have used it for quite some time).

As noted in GFI's article, "... [Bayesian] spam detection rates of over 99.7% can be achieved with a very low number of false positives!" However, even one false positive can be one too many. That's why I use the Quarantine plug-in in SpamPal to give me a plain-text version of every message that got marked as suspected spam (alas, there is no expiration of saved plain-text copies of the messages in the quarantine folder so I had to write a batch file to expire them, and which I gave to the Quarantine plug-in author to put on his site). I also move suspect messages (i.e., spam tagged) into my Junk folder in Outlook which has auto-archiving enabled to permanently delete messages within it that are over N days old (I usually set it to expire in 3 days although I'm tottering on just 1 day). They'll go away if I don't attend to them. Unfortunately auto-archiving or message expiration isn't available in other e-mail clients, like Outlook Express.

> *I've tried SpamPal and hated it. It tagged a lot of very legitimate
> e-mails ... including all that came from Road Runner accounts.*

Then don't use the overly aggressive and super vigilante DNSBLs (DNS blacklists), like SPEWS. If you use SPEWS, expect lots of false positives. I used to use SORBS but discontinued using that list when they choose to incorporate the SPEWS list (SpamPal indicates the interdependency by enabling SPEWS if you select SORBS). I've had 2 false positives by SpamPal since mid-February (whereas I get about 2 false positives per month via the ISP's server-side spam filter, so I may have to disable server-side spam filtering). I use the following blacklists: SpamHaus SBL+XBL, Composite Blocking List, ORDB, SpamCop, NJABL, and blitzed.org (although I might drop this last one). Use the Quarantine plug-in to capture the content of any spam-tagged e-mails but which are in plain-text format. If you're unwilling to configure the product to behave as YOU want regarding your e-mails experience then don't use that product. Same for anti-virus and anti-spyware products.

--

***** Post replies to newsgroup - Share with others *****
Email: lh_811newsATyahooDOTcom and append "=NEWS=" to Subject.

Re: Outlook should process rules first, before junk.