

Quoting entire messages (top posting) and HTML attachments are evil

A radical solution to get rid of them

[Introduction](#) [Download](#) [Operation](#) [Updates](#)

Introduction

I was really fed up of receiving mail messages which contain the same text once as plain text and once as HTML (I have just discovered that somebody sends message which even contain an attachment of "multipart/alternative" type, which contains two sub-attachments of which one is HTML ... there is no end to perversion). I was even more fed up of those guys who reply to a message of mine quoting my entire message at the end, instead of snipping only the parts they are really replying to (what is called top posting on Usenet, and heavily and rightfully criticized, and also against [netiquette](#)) :

bad manner	good manner
Dear Lucio, with reference to your message I have to say what follows.	Dear Lucio, with reference to your message I have to say what follows.
On 29 February 2000 Lucio Chiappetti wrote : > pinco panco blagulon exarzur otzhaxwq sobaka > pallino vitelli Bremsstrahlung notwithstanding	On 29 February 2000 Lucio Chiappetti wrote : > pinco panco blagulon exarzur otzhaxwq sobaka > pallino vitelli Bremsstrahlung notwithstanding
On this I fully agree while I disagree on the rest	On this I fully agree while I disagree on the rest
>txet emas eht niatnoc hcihw segassem liam >gniviecer fo pu def yllaer saw >ohw syug esoht fo pu def erom neve saw I >LMTH sa ecno dna txet nialp sa ecno >and so on and so forth ... >... for lines unnumbered >fo daetsni ,dne eht ta egassem eritne >ym gnitouq enim fo egassem a ot ylper	And therefore I do not quote it.

In particular a message in top-posting style reproducing my original message is a pain and a waste when archiving ... I have no need to archive again my original message which is already in my folder (and anyhow the message with the appendage is less legible). In the rare case I'd need to distribute the entire correspondence to others, a MIME digest is a much more efficient solution,

Some people use this signature to show how top-posting is silly:

```
--
Answer: Because it makes conversations flow in a nonsensical order.
Question: Why is top-posting wrong?
```

Therefore, after some attempts of dealing with this while archiving the mail message to a folder (I use [pine](#)), or using procmal, I finally devised this system (based on awk and integrated within my [procmal](#) rules.).

- if the message contains quoted text (i.e. if the body or first text/plain attachment contains lines starting with the standard quotation character (>) ...
- ... or if the message contains Content-type references to text/html
 - then I scan it and locate the start and end record of the unwanted parts.
 - an "improper quotation" is a sequence of lines prefixed with the standard quotation character, which are not part of a forwarded message (so far identified by -----Original Message -----) and are not followed by any non-blank text (excluding a standard signature)

- an "unwanted attachment" is a part of a MIME multipart message (as defined in [RFC 2046](#)) of type text/html. Currently the scan stage parses the start and end of any attachment, while the rejection is deferred to the following stage. I added provision to cope with the case an attachment is also a MIME multipart object, and descend recursively into it.
- Only if the scan has been tentatively successful, I submit the message to a filter, which passes the message through except for the unwanted parts, which are replaced by a placeholder.
 - **IMPROPER QUOTATION (TOP-POSTING) OF n LINES DELETED** for a bulky quotation
 - **HTML ATTACHMENT DELETED** in a one-line text/plain attachment, instead of an HTML attachment

Annoying behaviour by Outlook

I found that some correspondents using Outlook Express generate (in a way they are apparently unaware of) replies in a curious (and annoying way). I.e. the reply is not part of the normal body, but is disguised as a forwarded message (in turn, to make things trickier, it can be identified by ----- Original Message ----- or -----Original Message----- and the quote can or cannot be prefixed by the standard quotation character) !

Therefore I was compelled (in v1.3) to disable the pre-existing protection for forwarded messages in the particular case of replies generated by Outlook Express.

Download and installation

Since the above can be of general use I make available, with no guarantee or liability, what I have done according to the [GNU GPL](#).

Please follow these steps for installation :

- Insert in your `.procmailrc`, in the place most appropriate for it, a line calling my rule, e.g. `fi $PMDIR` is your procmail rule directory
`INCLUDERC = $PMDIR/rc.quote-html`
- Get and install in such directory the following file [rc.quote-html](#)
- Optionally customize it according to the comments in the first section of the file, namely :
 - set `PMDIR` if not already done
 - uncomment the generation of `TODAY`, an environment variable containing today's date, if you wish to keep dated backups. It is commented in my version since I set it elsewhere (I use it also for spam filtering).
 - set `BACKUPFOLDER` if you do not like my default. See below for backup operation.
- Get and install in the same procmail directory the analysis awk file [cleanbadquote.awk](#)
- Get and install in the same procmail directory the filter awk file [cleanbadquote.do.awk](#)
- Test it, sending some "bad manner" mail to yourself. If you do not want to test it on real mail, uncomment in [rc.quote-html](#) the line making reference to a Subject: TEST-TEST-TEST and send yourself mail messages with such subject. Comment it again when satisfied !
- If you have implemented dated backups and want to be perfect, arrange to clean up periodically the backup directory getting rid of older folders, inserting in a crontab something like this

```
find . -ctime $for -o -size $zero -type f -exec rm -f {} \;
```

which deletes the folders older than for days or of zero size (zero is not 0 but has to be found by trial, e.g. if your folder are managed by pine taking into account the "FOLDER INTERNAL DATA" used by pine).

Principles of operation

- All messages which contain lines starting with the standard quotation character `>` or which contain anywhere the string `Content-Type: text/html;` are submitted to the analysis script.
- Those which are not submitted, or which fail the script (for instance because the Content-type string occurs in the text, because the quotation character occurs not in the mail body, because the file has a MIME improper format) are flagged as `WELLBEHAVED=good` and are not subject to filtering.
- Otherwise the `WELLBEHAVED` variable contains the runstring to be used to call the filtering script. This is documented in the awk code.
- If a message is flagged as not well-behaved it is submitted to filtering.
- Before the actual filtering is done, a copy of the original mail is saved to a backup folder. In the default arrangement ("dated backup") a new folder is opened every day, with a name like `2006Oct24`, and is located in a hidden subdirectory `.A.Backup` of procmail mail area `$MAILDIR`. This way one can always access for some time the original message if something goes amiss in

the filtering.

- The filtered message is instead delivered wherever further procmail rules dictate. The fact it passed through the filtering program is indicated by the addition of the header keyword
X-LC-Rule: Possibly removed bad quoted text or unwanted (HTML) attachment
- The actual successful removal is indicated by the insertion of placeholder text (in lieu of an improper quote) or of a one-line text attachment (in lieu of an HTML attachment).

Update history

S/w V1.0 and web page established on 30 Oct 2006

S/w V1.1 (support to nested multipart attachments) on 06 Nov 2006

S/w V1.2 (support to Mac multipart attachments; boundaries not enclosed in quotes and content-type not immediately following boundary) on 19 Jan 2009

S/w V1.3 (handle top posting disguised as forward by Outlook) on 26 Aug 2010

Web page last updated on 26 Aug 10 11:54

Bugs

I suspect this software may not work well if the e-mail being filtered is encoded as "Quoted-Printable".

Possible improvements in customizing replacement messages (turning them off ?)

Possible improvements in supporting other quotation strings, and/or other kind of "bad attachments".

Contacts

I have no time to actively maintain this software other than for my own use. However I will be glad to receive communication of problems (or improvements) to it at my e-mail address `lucio` in domain `lambrate.inaf.it`.

[Lucio Chiappetti](#) - [IASF Milano](#) - [INAF](#)

