

Message Threading in E-mail Software

By Jacob Palme¹, e-mail jpalme@dsv.su.se, at the research group for CMC² (Computer Mediated Communication), which is a part of the K2Lab laboratory³ at the DSV⁴ University Department of Stockholm University⁵ and KTH Technical University⁶.

First version: 6 December 1998. Last revision: 26 December 1998.

Abstract: Threads are sets of messages, which are, directly or indirectly, replies to each other. This paper discusses different ways, in which various e-mail products support threads.

A thread is a set of messages, which are replies to each other. Since there can be more than one reply to a message, a thread is actually often a tree structure, as shown by figure 1:

Figure 1: An example of a message thread

Recognizing threads

E-mail messages have certain headers, which are used to recognize threads:

Message-ID

Specifies a globally unique identification of the current message.

In-Reply-To

May contain the Message-ID of the message, to which the current message is a reply. (Older mail software sometimes puts other, less useful information into this header.)

References

May contain a list of the Message-IDs of all the messages in the chain from the current message back to the start of the thread. If the thread is very long, this list may be abbreviated in the middle, but the first and the last message should always be present. (Older mail software uses this field to identify other messages, which the current messages refers to.)

Supersedes

May contain the Message-ID of a previous version of the current message. (Some

¹ <http://www.dsv.su.se/~jpalme>

² <http://www.dsv.su.se/~jpalme/DSV-CMC-Home-page.html>

³ <http://www.dsv.su.se/research/K2LAB/K2LAB.html>

⁴ <http://www.dsv.su.se/>

⁵ <http://www.su.se/>

⁶ <http://www.kth.se/>

software may use “Replaces” for the same purpose.)

Look at figure 1:

Figure 1 again: An example of a message thread

In this example, message B and E have the Message-ID of Message A in the In-Reply-To and/or the References header.

Message C and D have the Message-ID of Message B in the In-Reply-To header, and have the Message-ID of Message A and Message B in the References header.

To be able to use these headers, mail software sometimes contains a data base, which, given a Message-ID, returns the message with this Message-ID. Thus, given an In-Reply-To header, the message referenced in this header is returned. Note that since the Message-ID is globally unique, such a data base can be used to copy the link from the mailbox data base of the sender to the mailbox data base of the recipients of the messages.

Threads can also be recognized in a very different way, by using the Subject field. Replies to a message often have the same Subject value as earlier messages in the thread, except that the four characters “Re: ” are usually added to replies. “Re: ” is not added to Subjects which already start with “Re: ”. (Sometimes, although this is not a standard, “Re: ” in other languages than English is used, for example “An: ” which is an abbreviation of the German word for reply, “Antwort”. This method should be avoided, since it causes technical problems with international messages. “Re:” is actually an abbreviation of a Latin word, “referrer” and

Latin is an international language. Message systems can, if they so prefer, show “Re: ” as “An: ” in the user interface, but in the protocol on the wire, the value should always be “Re: ”.) Sometimes, although this is not a standard, “Fwd: ” is added to the Subject when forwarding a message. Thus, by finding all messages with the same Subject, after stripping “Re: ” and “Fwd: ”, will find all messages in a thread.

However, a person writing a reply can change the Subject. In such a case, the In-Reply-To header might link this message to a thread even though the Subject seems to start a new thread. Some people think this is an advantage, since a change of the subject might indicate a new topic. Sometimes, to keep the thread while changing the subject, the old subject is kept, with “(was)” in front of it. Thus, if a message has the subject “Train schedules (was) Bus schedules”, this shows that this message belongs to the same thread as earlier messages with “Bus schedules” as the subject.

Note that the set of messages which are recognized as threads may differ with these two methods, since the header fields Message-ID, In-Reply-To, References, Supersedes and Subject do not always contain the necessary information in the way described above.

Can threads be merged?

Figure 2: Merging threads

One message may have multiple values in the In-Reply-To, References and Supersedes headers (an older version of the e-mail

standard only allowed a single value in the In-Reply-To header). Suppose you have the structure in figure 2. Will then the addition of Message G cause two previous threads to be merged into one thread? Or does Message G belong to two different threads? Different mail clients handle this in different ways. Some clients will only recognize a thread by the first value in these headers, so that Message G only belongs to the left, and not to the right, thread in Figure 2.

User commands to follow threads

Clickable links

Mail systems usually show the In-Reply-To, References and Supersedes headers. The reader of a message can then, with some message systems, click on these headers to get to earlier messages in a thread.

Even though there is no explicit link from a message to its replies, a mail system can create and show such reverse links, so that a reader can click on links both to follow the thread back and forward. Such reverse links are useful, because they allow a reader of a message to check if there are already replies, before writing his or her own reply.

Below is an example of how this can be handled in an e-mail software.

Date: Thu Aug 27 14:35:19 1998

Author: Mary Smith (13)

In-Reply-To: London meeting <-The user can click here to get the previous message in the thread

To: Meeting planning

Language: English, Swedish <- The user can click here to get a Swedish translation of this message

Re: London meeting

I suggest that we meet at Trafalgar Square on 8 p.m. the night before the meeting.

The user can click on the icon or the subject below, to see the replies to this message:

Re: London meeting , by John Clarke Thu Aug 27 16:18:54 1998

Re: Trafalgar Square , by Jacob Palme Sun Sep 06 20:55:03 1998

Next Message Figure 3: Clickable links in the KOM 2000 system

Clickable links allow a user to go one step at a time up and down a thread. Some message systems have commands to show a whole thread. There can, for example, be a command to get a list of all the messages in a thread, or a command to find the chronologically first or last message in a thread. Or a command to scan all messages in a thread upwards and downwards.

Another variant is commands to sort the messages in a mailbox by Subject (disregarding "Re: " and similar prefixes).

Note an important difference between these two methods. The first method will recognize a thread, even if the thread starts in one mailbox and moves to another mailbox. The second method will only show messages in a thread within one mailbox.

There are several reasons why threads can move between mailboxes. One cause of this is replies to a mailing list message, sent only to the author of one of the messages. Incoming messages may be sorted into a separate mailbox for each mailing lists, while personal messages may be put into the standard “In” mailbox.

Listing messages in threads

Some mail system will sort messages according to their place in a thread structure. A listing of messages might thus look like this: London meeting, by Mary Smith 20/12/98 15:15 Re: London meeting, by John Clarke 22/12/98 15:23 Re: London meeting, by Syd Gray,23/12/98 08:13 Re: London meeting, by Dan May 23/12/98 16:30 Stockholm meeting, by Fred Sterling, 22/12/98 16:23 Video meeting, by Tom Sitler, 27/12/98 16:39

Figure 4: Listings with indentations to show thread structure

(example from Web4Groups)

A variant of this is to show only two levels. The first message in each thread is not indented, but all other messages are indented. Example:

London meeting, by Mary Smith 20/12/98 15:15 Re: London meeting, by John Clarke 22/12/98 15:23 Re: London meeting, by Mary Smith,23/12/98 08:13 Re: London meeting, by Dan May 23/12/98 16:30 Stockholm meeting, by Fred Sterling, 22/12/98 16:23 Video meeting, by Tom Sitler, 27/12/98 16:39

Figure 5: Showing threads with only one level of indentation

A third variant is to list only one message for each thread. A user might then click on this message, to get to a list of the messages in this thread, either in a new window, or in an expansion of the current window. Example:

London meeting, by Mary Smith 20/12/98 15:15 Re: London meeting, by John Clarke 22/12/98 15:23 Stockholm meeting, by Fred Sterling, 22/12/98 16:23 Video meeting, by Tom Sitler, 27/12/98 16:39

Figure 6: Expansion of a thread as requested by the user

In the user interface in Figure 6, the symbol indicates a closed thread, and clicking on this symbol changes it to while the thread below it is opened.

Note that the first of these methods, to show the full tree structure by indentations, will not work well if the user requests a list of only unseen messages. The other methods, however, will work as well for lists of unseen messages as for lists of all messages.

Another advantage with only showing one level in hierarchical listings is that users do not always put replies in the right place in the tree. For example, many users write replies on the latest message in a thread, even if their message is in reality more of a reply to an earlier message in the thread.

Special handling of Supersedes

Since most users only want to see the most current version of a message, mail systems will often only show the last message in a Supersedes chain of messages. However, some systems have commands for a user to

see earlier messages in this chain, if they explicitly ask for them.

Risks with incorrect implementation of Supersedes

Links must refer to the right version

If a message B has an In-Reply-To link to a message A, and if A gets Superseded by C, then clicking on the In-Reply-To link should get message A, not message C. This is necessary, see the following example:

Message A: Do you like Jesus? Message B: (In-Reply-to A) Yes, he is a good person Message C: (Supersedes A) Do you like Saddam Hussein? If clicking on the In-Reply-To link in B would show message C, then it would seem as if the author of message C liked Saddam Hussein!

Do not physically delete in other people's mailboxes

There is a risk with Supersedes, if it is implemented to really delete the earlier messages. Many users would not be happy with having messages deleted from their mailboxes. Because of this, Supersedes is often implemented so that all message versions are available. Users want Supersedes, but some implementors are wary of supporting it because of this risk.

References

D. Crocker 1982: Standard for the format of ARPA Internet text messages. STD 11, RFC 822, August 1982, at URL [D. Crocker: at URL: ftp://sunic.sunet.se/rfc/rfc2184.txt](ftp://sunic.sunet.se/rfc/rfc2184.txt)
Horton, M.R. and Adams, R 1987: Standard for interchange of USENET messages, RFC 1036, at URL

<ftp://sunic.sunet.se/rfc/rfc1036.txt> Palme, J. and Tholerus, T. 1992: SuperKOM - Design considerations for a distributed, highly structured computer conferencing system, in Computer Communications, vol. 15, no. 8, October 1992 pp 509-518 Available in PDF (Adobe Acrobat) at URL:

<http://www.dsv.su.se/~jpalme/w4g/superkom-design-cons.pdf> Palme, J., 1995: Electronic Mail, Artech House Publishers, London-Boston ISDN 0-89006-802-X, at URL <http://www.dsv.su.se/~jpalme/e-mail-book/e-mail-book.html> Palme, J. 1997: Common Internet Message Headers, RFC 2076, at URL <ftp://sunic.sunet.se/rfc/rfc2076.txt> Palme, J. 1998: A Proposal for Extending Eudora with Thread Support, at URL <http://www.dsv.su.se/~jpalme/ietf/thread-support-proposal.html>

Other documents of interest