

Review Strategies Utilizing Near Duplicate and Email Threading Technology

Rachi Messing, VP Customer Solutions – Equivio

The usage of Near duplicate and Email Threading technology has become relatively common in large discovery document databases. A common question that is asked by Litigation Support and Case Management staff is “Which documents actually need to be reviewed once the results of the technology are available?” Overall, it is important to remember that the only reduction of review data through the Near Duplicates application should be from the identification of duplicate documents. Otherwise Near Duplicate analysis is about enabling a more efficient review of the data, not reduction of the review population. Email Threads can however significantly reduce the volume of documents that need to be reviewed when deployed properly.

There are 3 categories of documents that require review after running Near Duplicate/Email Threads:

1. Inclusive emails – the email in the thread that is identified to fully contain the content of other emails in the same thread
2. Unique attachments
3. Unique standalone documents

To achieve maximum efficiency, only a single, unique copy of any md5hash version of a document (whether attachment or standalone doc) should be reviewed. Once reviewed the decision made for the unique document should be propagated to all other copies of the document. There are some people who are not comfortable with this approach and require a copy of the attachment to be reviewed in each thread where it exists in order to see the attachment in the context where it was sent.

Once documents that require review have been identified, the next question that arises is typically “Now what?” The review teams require guidance on how to go about reviewing the documents and applying their decisions in order to gain the most efficiency possible. Below are some suggested guidelines to help a reviewer or case administrator determine the most efficient workflow that will work for them.

If the Inclusive email is Not Responsive, then there is no need to review the rest of the email thread as that content is going to be Not Responsive as well. The reviewer can mass tag the rest of the thread as Not Responsive but should remember to still check all attachments needing review.

If the Inclusive email is Responsive, then one must determine their review and production strategy. Content that exists earlier in the thread will be produced and seen through this last Inclusive email. One may decide to take the approach to mass tag the entire thread as Responsive. This very well may result in the production of emails that on their own are Not Responsive; i.e. over production. However, this has the potential to drastically reduce the review time. This is something that counsel may want to discuss with the opposing party. If the other party has an issue with the over production, then an alternative to offer them is the production of only Inclusive emails (plus all attachments from anywhere in the thread).

One challenge that is sometimes raised with producing only the Inclusive emails is that one may potentially not see all of the sender and recipient information (To/From/CC/BCC) from the entire thread. It is possible that the participants have changed throughout the thread and all are not present on the final Inclusive email. If this is a concern, then a separate Participants field can be populated that identifies every unique email participant throughout the entire thread. This field can be used for search and production. The Inclusive email can be examined to determine where in the thread a specific address appeared.

If the Inclusive email is determined to be Privilege, then the review team must first determine if the entire thread is privilege. If yes, then they can mass tag the entire thread as privilege. If the entire thread is not privilege, then the first step is to determine where within the thread the privilege material exists. If the earlier part of the thread is not privilege and the privilege content is only in the later part(s), then one must determine their review strategy. In some cases it may make sense to tag the entire thread as Potentially Privilege and have a higher level review team determine the exact point in the thread where the split between the privilege and non-privilege content exists. Others take the approach to review each email within the thread separately and make a privilege determination for each individual part of the thread. This eliminates the use of mass tagging entirely. If an agreement was reached to only produce Inclusive emails, then either this entire thread will need to be withheld as privilege, the Inclusive email will require redaction, or an exception to the policy will need to be made. If the latter option is chosen, then the part of the thread immediately prior to the start of the privilege information will need to be produced.

If the earlier part of the thread contains the privilege content but then at some point the privilege was broken, a review and production strategy question needs to once again be addressed. If the approach to these emails will be to redact the privilege content within a Responsive email, then the Inclusive email should be tagged as Privilege-Redact. If the production decision is to produce more than just the Inclusive emails, then the rest of the thread should be mass tagged the same way up until the point where the privilege broke. All emails prior to that point should then be tagged privilege.

If the approach will be to not redact the privilege content, then an interesting decision needs to be made. The content of the privilege email will be available to the other side within this Inclusive non-privilege email. Thus, there is no real risk in producing the earlier emails in regards to accidental disclosure. Therefore, just as the Inclusive email is going to be tagged Responsive the rest of the thread can be mass tagged as Responsive as well, saving the most review time. However, there are some who have concerns with this approach. The result will be that an email that is technically determined to be privilege on its own (the earlier email in the thread) is going to be produced even though the content of this privilege email is being produced later via the non-privilege Inclusive. While this approach may increase efficiency and streamline the workflow, it may give the opposing side reason to question the overall approach to the privilege review.

Use of Near Duplicate and Email Threading technologies can add efficiency and organization to a review. With the proper procedures in place throughout the review, the end user can rest assured about the overall defensibility of the approach as well.