



White Paper

Content-Centric E-Mail Message Analysis in
Litigation Document Reviews

Introduction

E-mail messages have become the primary focal point in many, if not most, litigation document review projects. First, e-mail materials increasingly make up the bulk of the documents exchanged in fact discovery. Second, and equally important, e-mail is a rich source of evidence, often in the form of plain talk that differs from official corporate memoranda prepared at the same time. E-mail conversations also provide key anchors for refreshing the memories of fact witnesses, who often cannot clearly remember specific events and conversations but who can authenticate e-mail messages they wrote. Given how long e-mail messages have been considered critical evidence, the inefficiency of most e-mail review projects is surprising, even after years of experience. However, new technology offers new ways in which to extract maximum information from e-mails with the least amount of pain.

Organization is the Key to Efficient E-Mail Review

Document reviews are driven by two basic technologies. First, some type of search is used to identify materials that require further analysis or categorization. Second, organization prepares the search results for efficient review. A variety of methodologies can be used to accomplish either of these steps, though different approaches impact the time and expense that will be required to achieve document review goals. Organization, in particular, has received somewhat less attention than search, in large part because the technology requires adjusting review workflow—a more complicated process than using a different search paradigm to identify documents to feed into a review.

Most traditional e-mail reviews are organized around the way that the e-mail archives have been collected by legal teams. Most often, this means that the e-mail messages are segregated by the custodian or computer from which they were collected. However, when review is conducted purely on a custodian basis, reviewers may miss significant pieces of the total story because their analysis is limited to an incomplete archive of all possible documents. In addition, because the archives are stored in separate information silos, it can be difficult to successfully combine information gathered through separate reviews of custodian documents into a single, seamless and comprehensive story.

Reviewing e-mail by custodian is also inefficient. At the same time that it dilutes the power of the story, custodian review also involves maintaining duplicate copies of messages that appear in multiple custodian e-mail archives. Thus, when e-mail is separated and reviewed by custodian, the same e-mail message will be reviewed once for each mailbox in which a copy is stored. The amount of e-mail overlap between professional colleagues can be substantial.

Moreover, as is typical in e-mail conversations, prior messages appear repeatedly in the downstream e-mail chain. This is the problem of “e-mail containment”. E-mail containment generates significant redundancy and inefficiency in the review process, and is a source of considerable frustration for review lawyers.

A somewhat more efficient approach for reviewing e-mail combines e-mail messages from all custodians and sorts them by specific criteria, most typically date. This method permits global message de-duplication, so that the review team can choose to review only one copy of each unique e-mail message. In addition, when messages are sorted chronologically, the review team sees all conversations taking place, laid out in real time. This increases the likelihood that reviewers will identify subjective relationships between seemingly unrelated e-mail messages that are not linked by common subject lines or common authors and recipients.

However, a purely chronological approach to reviewing e-mail still makes it difficult to build a comprehensive picture of the evidence. E-mail conversations often extend over hours, if not days, and viewing messages in pure chronological order interrupts a reviewer’s ability to see the entire conversation from start to finish. Instead, one or more reviewers may see segments of the conversation, which may or may not include sufficient context from earlier communications. Incomplete e-mail thread segments may provide enough information to permit basic triage, but key messages may still remain unconnected because of the separation between them caused by the passage of time. E-mail messages can also be sorted by subject line and by date to keep threaded conversations together. However, this technique cannot prevent irrelevant messages due to “topic creep” from taking up reviewer time, nor can this method link related messages that have spun off under modified or different subject lines. This technique is also challenged by the difficulties of normalizing date and time information across different time zones.

Organizing E-Mail Messages by Related Content Increases Review Quality

E-mail message metadata such as date, author, recipient, and subject line provides important context for understanding the significance of messages. However, it is the substantive content of messages, not their metadata, that defines their significance—or irrelevance. Message text is the key information that links e-mail messages across wandering subject lines and distribution. New technology enables the analysis of e-mail message content to reconstruct e-mail threads, permitting reviewers to view e-mail messages in the context of an e-mail conversation, while gaining a competitive advantage over other review teams that use slower methods to learn the same information.

Organizing messages by content differs from finding duplicate messages, something that's been possible for some time. Related messages share varying degrees of commonality, typically through quoted portions of earlier e-mail messages to which they are responding. Technology now facilitates the isolation of this recurring content, and also uses it to identify relationships between e-mail messages.

Recurring content identifies relationships between e-mail messages that metadata analysis cannot find. For example, matching recurring content finds the relationship between two e-mail messages even when one of them is forwarding the prior message to entirely new recipients and bears a different subject line.

Tracking recurring content also permits better organization of e-mail conversations that split into distinct threads, either through topic drift or through follow-up messages with fewer or different message recipients. E-mail groupings can be organized for increased granularity to segregate such sub-threads from the larger conversation. Again, this permits more coherent review of potentially voluminous e-mail archives.

Grouping E-Mail Messages by Content Increases Efficiency and Reduces Costs

Once groupings of related e-mail messages have been identified, these messages must still be organized for efficient review. Efficiency goals include gaining a comprehensive understanding of the review materials and completing this task with the least amount of effort.

The content-centric groupings use a focal point to reduce review time. The fact that so many e-mail messages quote prior e-mail messages means that, in virtually all cases, only the last message in a thread needs to be reviewed. The last e-mail in a thread typically contains all the content from the foregoing messages. A content-centric approach to building the e-mail threads can analyze the thread to verify that the last e-mail does in fact contain all the content from the upstream message chain. This approach can increase a reviewer's effective rate of review by an order of magnitude and without overlooking any unique information.

Different review platforms offer additional tools for leveraging the analysis generated by content-centric grouping of e-mail messages. For example, the use of redlining to highlight the unique content in e-mail messages enhances review speed because non-unique message content can be disregarded once it has been reviewed the first time.

Conclusion

Grouping e-mail messages based on message text, rather than traditional metadata information, is a strategy that legal teams should strongly consider when organizing document reviews. Though it is a relatively new approach, benefits may be striking. First, because related messages will be reviewed by a single reviewer, the resulting analysis is more comprehensive than traditional reviews that scatter related information across multiple review batches and reviewers. Second, the reconstruction of the e-mail thread based on content dramatically increases review speed because the reviewer can focus exclusively on the last e-mail message in a thread. Content-centric analysis of the thread ensures that the last e-mail message does in fact quote all the previous layers of the conversation. Using this approach, legal teams can achieve the best of both worlds—better information at a lower total cost.

Warwick Sharp

Vice President, Marketing and Business Development at Equivio