Near-Duplicates and Email Threads in Equivio Zoom

Eliminating Redundant Data in Document Review
BOOSTING REVIEW PRODUCTIVITY AND REDUCING LITIGATION COSTS

Near-Duplicates and Email Threads are fully integrated components within Zoom, Equivio’s e-discovery platform for predictive coding and analytics. By grouping similar documents, users gain a decisive advantage – less time wasted reviewing redundant data, and more time spent analyzing the unique information critical to the case. Equivio’s proven, best-of-breed Near-Duplicates and Email Threads applications are used by hundreds of corporations, law firms and e-discovery service providers world-wide.

Litigating organizations are challenged by exploding document volumes and skyrocketing e-discovery costs. The organizations involved in litigation -- from corporations to law firms and government -- are searching for effective ways to cope with the document surge.

Zoom’s Near-Duplicates and Email Threads applications expedite the management and review of unstructured document repositories. The grouping of near-duplicates and email threads allows similar documents to be handled and treated together. The result: Equivio users slash the time and cost of document review, while ensuring the consistent treatment of similar documents.

The potential savings through the use of near-duplicate grouping and email threading are compelling. Near-duplicate email thread grouping consistently drives review cost savings of 30 to 50%. This translates into huge potential savings for users of the integrated Zoom platform.

THE DE FACTO INDUSTRY STANDARD FOR MANAGING DATA REDUNDANCY

Equivio’s Near Duplicates and Email Threads applications have become the de facto industry standard, having been used to streamline litigation review in thousands of cases over the last six years:

Proven: Equivio’s Near-Duplicates and Email Threads applications are used by over 80 e-discovery service providers, and serve the Department of Justice, the FTC, and hundreds of law firms and corporations. System throughput and scalability have been demonstrated in live mega-cases of more than 10 million documents.

Defensible: Equivio algorithms analyze content to verify that near-duplicate sets and email threads contain zero false positives. The elimination of false positives is critical in litigation review – otherwise a reviewer may skip a document, assuming it’s a near-duplicate, when in fact it is a totally different document containing key evidence.
Setting the standard: The near-duplicate algorithm detects near-duplicates through the entire range of similarity thresholds – 40% through 100%. Most near-duplicates have resemblance levels lower than 90%, which makes them difficult to detect by most available tools. Email Threads analyzes content rather than depending on unreliable metadata and supports any email format including plain text.

NEAR-DUPLICATES

The Near-Duplicates application detects and groups very similar documents - that is, documents with minor differences (e.g., contract versions containing a few different words). This set-centric approach enables replacement of linear review in litigation review – instead of reading document-by-document, reviewers can work set-by-set resulting in a more efficient and coherent review process.

Near-Duplicates identifies near-duplicates and organizes them into sets for efficient and coherent handling. This accelerates the document review process:

- **Stage 1**: Assign the near-duplicate sets to reviewers, ensuring that one reviewer handles a document and its near-duplicates in a systematic manner.

- **Stage 2**: Start review of a near-duplicate set with the pivot document, the most representative document in the near-duplicate set. After reading the pivot, the
reviewer decides whether the rest of the documents in the near-duplicate set can be skipped or require granular review. In most sets, the reviewer can apply a tag - e.g. responsive, not-responsive, privileged, etc. -- to the entire set based on the pivot.

- **Stage 3:** In some cases, granular review and individual decision on each document is required. Using the built-in Compare facility, the user zooms in on the differences in each near-duplicate vis-à-vis the pivot document. As a result, the user needs to read only the differences (which may be only a few words) rather than the entire document. This helps to ensure that the user does not miss minor, yet potentially critical, differences.

- **Stage 4:** The user can bulk handle a near-duplicate set based on tags. The Equivio groupings facilitate the consistent treatment of like documents -- for example, when marking documents as privileged or responsive.

![Unstructured document set](image1.png) ![Groups the near-duplicates](image2.png) ![Focus on the differences](image3.png)  
*Figure 1 Using Near-Duplicate Sets*

**EMAIL THREADS**

Zoom’s Email Threads application captures and reconstructs email conversations so that emails can be reviewed within their original thread context. By identifying the unique emails in a collection, the tool drastically reduces the number of emails that need to be reviewed.

The Equivio tool simplifies and streamlines the review of emails:

- **Stage 1:** Assign email threads for review, ensuring coherent review of email data.

- **Stage 2:** Start review of an email thread by reading the first “Inclusive”. An “Inclusive” is the last email in a thread. Equivio verifies that each Inclusive contains the entire history of the thread.

- **Stage 3:** Review the other Inclusives in the thread.

- **Stage 4:** Users can bulk handle all the emails in a thread, reducing time and cost, while also ensuring that all the emails in the thread are treated consistently.
Equivio’s content-centric technology for email threads allows users to review only emails that contain unique content. Applied to data in the Enron case, Equivio was able to reduce the overall review effort by 60%. The content-centric approach is robust, eliminating any dependence on the idiosyncrasies of email metadata.

**BOTTOM LINE: LESS TIME, LOWER COSTS, FASTER RESULTS**

By grouping near-duplicate documents and emails prior to document review, Equivio generates immediate, concrete benefits for users of the Zoom platform:

- **Reduced costs.** Reviewers are directed to what's important – i.e., to documents that contain unique information, and to the unique information within each document. Proven in hundreds of cases, Equivio’s technology consistently reduces review and handling costs by 30 to 50%.

- **Less time.** In many situations, it’s impossible to review all the documents in the given time window. Equivio enables prioritized review by zooming in on value-added, unique data.

- **Less risk.** By directing reviewers to the unique documents, and to the unique data in those documents, Equivio reduces the risk of missing critical information.

- **Consistent treatment.** The Equivio groupings allow reviews to apply tags across a near-duplicate set or email thread to ensure very similar documents are treated consistently. The need for consistency applies throughout the data cycle, from the feed of documents into the retention archive, through the implementation of storage policies, and the treatment of documents in a specific litigation or regulatory event.

- **Accessing the data users need.** By allowing the virtual suppression of redundant data, Equivio helps users cut directly to the information they need.
ABOUT EQUIVIO ZOOM

Equivio Zoom is an integrated platform for e-discovery analytics and predictive coding. Zoom brings together Equivio's proven technologies for e-discovery analytics in a unified web-based platform. Combining Equivio's best-of-breed near-duplicates, email threads and relevance components together with data import, ECA and enriched analytics capabilities, Zoom provides the tools you need for easier and smarter e-discovery.

ABOUT EQUIVIO

Equivio develops text analysis software for e-discovery. Users include the DoJ, the FTC, KPMG, Deloitte, plus hundreds of law firms and corporations. Equivio offers Zoom, an integrated web platform for analytics and predictive coding. Zoom organizes collections of documents in meaningful ways. So you can zoom right in and find out what's interesting, notable and unique. Request a demo at info@equivio.com or visit us at www.equivio.com.

Zoom in. Find out.