Practice Area
briefings

sponsored content graciously presented by McDermott Discovery

Predictive Coding at McDermott Discovery: Real Experience, Real Results

By Alison Silverstein & Martha Louks

There has been a lot of discussion among e-discovery professionals about how Technology Assisted Review (TAR) can be used to improve the efficiency and reduce the cost of document review. Typically, TAR is thought of exclusively as "predictive coding". In reality, predictive coding is just one way technology can assist a document review project and make that project more cost-effective and efficient. Though each matter will require its own unique approach, there are a few McDermott Discovery techniques that are proven to reduce the burden and risks while increasing the rate of review and accuracy.

Predictive Coding
In a typical McDermott Discovery predictive coding workflow, a relatively small number of "seed"

or sample documents are evaluated for responsiveness by a core team of attorney reviewers. The system uses these decisions to code all documents in the data set. Additional seed documents are then reviewed by the core team, and the system is able to refine its understanding of the rules of responsiveness and relevance. A control set is used to generate benchmark statistics, including recall and precision, which allow the users assess the accuracy of the exercise. When an acceptable percentage of responsive documents have been found by the system (recall) and the accuracy of the coding is sufficient (precision), the training process is concluded. The threshold for recall and precision varies from matter to matter and is often determined by the legal team or through agreement with the opposing party or government agency.

At the conclusion of the training process, the system will attempt to categorize each document as

responsive/relevant or non-responsive/non-relevant. Depending on the technology being used, some documents will remain uncategorized. At this point, the legal team is able to decide whether to review the remaining uncategorized documents. Often, the legal team will also choose to review the responsive set prior to production, as a final means to ensure that irrelevant information is being produced and sometimes for no other reason than to know what will be produced before it goes out the door.

McDermott Discovery has successfully used predictive coding workflows in several cases to eliminate large swaths of documents from linear review. Making a small up-front investment in the review of seed documents yields enormous cost savings because it segregates and eliminates the need for attorney review of the non-responsive data. The number of seed documents reviewed usually consists of a very small percentage of the

total universe of client data, yet the non-responsive data eliminated from the review process is often extremely large. In general, review of approximately 1-2% of the dataset will lead to a 70-80% reduction in the data requiring review.

Reducing the percentage of documents requiring review leads to a commensurate reduction in the cost of the review: eliminating 75% of the document universe reduces the review cost by 75%.

Other Uses of Technology to Improve Review Efficiency and Reduce Cost

While predictive coding is very often an appropriate and effective solution, there are times when it is not. In these situations, there is alternative technology that can be employed to increase the rate at which reviewers move through documents, improve the accuracy and quality of the decision-making, and reduce review costs. Email threading and concept clustering are two examples.

Email Threading

Email threading identifies the relationship of emails in a document set to each other, making it possible to group emails by conversation. Additionally, email threading identifies the most inclusive email in a thread, as well as emails that contain duplicative content but could not be de-duplicated during data processing due to minor text differences.

Workflow approaches to leveraging email threading technology will naturally vary from matter to matter, but we have found that the following approaches have been effective in increasing the rate of review and reducing costs:

- *Group Review Batches by Conversation.* Sorting the documents in each review batch by email conversation, such that the most inclusive email occurs first in the batch, will typically double the rate of review, while resulting in better and more consistent decisions by the reviewer.
- *Reviewing Email Strings.* Sometimes legal teams choose to review only the last email in a conversation rather than what otherwise would be multiple emails involving overlapping content. The efficiencies gained by the assignment of all emails in a string to a single reviewer and the elimination of all of the component emails of the larger email string often results in time and data reductions of between 15% - 30%.
- *Quality Control.* In addition to speeding up the process and eliminating large chunks of data to be reviewed, email threading technology can be used to monitor and indeed improve the quality of the review project. For instance, if an email is designated as privileged, it may be necessary to designate all emails in the conversation as privileged. Email threading technology will allow you to quickly and effectively identify other emails in the conversation to ensure consistency in privilege calls.

Concept Clustering

In a clustering workflow, documents that are conceptually related are grouped together. Review batches can be organized so reviewers are seeing similar documents consecutively in their batches. By concentrating similar documents together, reviewers can more quickly become familiar with a particular subject matter or type of document. Their improved understanding of the data set will enable them to make faster, more accurate, and more consistent decisions about the documents. McDermott Discovery has had great success accelerating the rate of review using concept clusters. In most cases, we typically see a 60-80% increase in the decisions per hour made by each attorney reviewer.

Conclusion

There is no "one size fits all" technology for use in managed review projects. However, predictive coding, email threading and concept clustering technology, in addition to the plethora of other available technologies, can be game-changers when used the right way in the appropriate cases. At McDermott Discovery, we use technology to reduce the volume to be reviewed, increase the speed by which our attorneys review what remains, and improve the overall accuracy of our decisions. Better quality at a substantially reduced price. We do discovery differently. We do it right. **PAB**