



Active Learning – Prioritised Review

The Challenge

Our client received a tranche of disclosure from another party totaling around 55,000 documents. Their aim was to review all 55,000 documents, but to prioritise the review to find documents similar in nature to their own key documents first.

Solution

Relativity has various tools to identify similar documents. Documents can be grouped by near textual duplicates, clustered by concepts or fed into a Technology Assisted Review (TAR) workflow.

In this matter where the client wanted to review all documents, but prioritise the order of review, the best solution was to deploy Relativity’s active learning workflow.

Active learning learns from decisions made on documents as they are coded in the review workflow. It uses these decisions to continuously deliver documents to a review queue based on what the software believes to be the next most similar document to those already coded as relevant.

As the review progresses, the review queue is continually updated based on the decisions being made.

In this matter we were able to take the documents that the client had already identified as key documents in their own disclosure and use these as a pre-coded set to train the system and kick start the active learning review queue.

Outcome

The active learning project showed its value from the start. Using only a very small seed set of around 260 key documents, the project returned a mean relevance rate of 45% for the first 2,000 documents. By the second 2,000 documents, the mean relevance rate had fallen to 24% and by the third set of 2,000 documents the mean rate had fallen to just 12%.

In overall terms, the mean relevance rate fell from an initial peak of 63% (for the first 200 documents) to 6.8% by the end of the project, with the last 20,000 documents having a mean rate of just 1.54%.

The below graph illustrates how using active learning enabled the review team to review the most relevant documents sooner.

Overall Mean Relevance rate





Benefits

If this was a traditional active learning project, where the client was reviewing their own documents for disclosure, we may well have stopped the project after the first 15,000 documents to test the results. At this point the relevance rate for each review set of 200 documents had dropped below 5%. The 5% rate is a guide only and a project can be stopped and tested at any point to test the results and determine whether it is proportional and reasonable to continue to review the remaining documents.

If we assume that, as part of the testing, the team reviewed another 1,000 documents to assist in the validation of the results, the client would have reviewed a total of 16,000 documents, leaving 39,000 documents unreviewed.

Applying this logic to the final number of documents actually reviewed in this matter, there would have been 25,000 documents left to review (excluding documents which were suppressed as duplicates).

To review 25,000 documents would normally take one reviewer approximately 62.5 days (based on 400 documents per day), or 468.75 hrs (based on 7.5 hrs per day) to review. If we assume that the reviewer was charged out at the rate of \$250 an hour, then the cost to review the final 25,000 documents would have been \$117,187.50.

From the actual project we know that in the final 25,000 documents the review team only coded 247 of these as relevant. This meant that each additional relevant document

would have potentially cost \$475 to locate applying a traditional linear review model. If at the start of the matter the cost to review these 25,000 documents had been calculated using that figure, the estimated cost for first level review alone would have come to \$26,094,382.

Clearly that is an exaggerated estimate, to illustrate the disproportionality of reviewing the last 25,000 documents in order to locate only 247 relevant documents, but by not reviewing the last 25,000 documents the client however could still have theoretically saved the \$117,187.50 in review costs.

Whilst these savings were not available to the client in this matter as they always intended to review all 55,000 documents, there were still savings to be made. In suppressing close to 15,000 documents as duplicates, applying the active learning workflow meant that the client only ended up reviewing just over 41,000 documents. The 5% rate is a guide only and a project can be stopped and tested at any point to test the results. The tests are used to determine whether it is proportional and reasonable to continue to review the remaining documents.

Your Project

Ask about how we can deploy active learning in conjunction with our Managed Document Review team to reduce costs and speed up your review even further.

About Law In Order

Established in 1999, Law In Order is the leading supplier of end to end document and digital solutions to the legal industry providing expert litigation support through our cost-effective document production, expert discovery management and specialist court services. Law In Order operates 365 days a year and has offices in Brisbane, Sydney, Melbourne, Perth and India. This ensures that our clients have unrestricted access to the services they require and are able to contact experienced litigation support professionals for assistance and advice at any time.