Continuous Active Review Cuts Cost by Over 85%

“Catalyst’s review platform and its ability to support Continuous Active Learning enabled our client to cut the time and cost of review by over 85%.”

**Client Snapshot:** Japanese Multinational
- International patent dispute
- 3.6 million Japanese and English docs
- Catalyst cuts review costs by millions

© 2015 Catalyst Repository Systems. All Rights Reserved.
Our client was a multinational Japanese company facing a large document production in an international patent dispute. The initial review collection exceeded 2 million documents. After a series of rolling uploads, which continued throughout the review, the population slated for review grew to 3.6 million. Facing millions in review costs, the client sought an alternative to linear review.

Review time was short. The client’s goal was to finish the review in four weeks with a small team handling the project. The documents were primarily in Japanese, with some English in the mix, and many involved highly technical subject matter.

**Estimating Richness and Training**

Even though the client had taken steps to remove junk and other documents not subject to production, the collection's estimated richness was still miniscule. An initial systematic random sample of 1,000 documents (97% confidence with a 3.5% margin of error) suggested that there were fewer than six relevant documents in every thousand that might be presented through a linear review. As is often the case in litigation, richness was low at 0.6%.

Before Catalyst was engaged, a team of lawyers had reviewed about 10,000 documents found through keyword search. For many TAR 1.0 engines, which have a limited training phase, these judgments would have been of no use. Because Catalyst’s technology, Insight Predict, is a TAR 2.0 engine that uses continuous learning and continuous ranking, we could make use of these judgments as initial training seeds.

As you can see from the yield curve on the next page, the initial training using the 10,000 seeds proved effective. It indicated that almost all of the relevant documents could be found after reviewing just 17% of the total review population. This meant that the review team could immediately exclude most of the non-relevant documents and start finding relevant documents many times faster than the day before. There was no need for the team to spend non-productive hours looking at largely irrelevant files selected randomly for initial training.
Optimizing Review with Continuous Active Learning

The initial training worked. Richness in the documents presented to the review team jumped from 0.6% to as much as 35%, which represented a 60-fold improvement in review efficiency. At the same time, the reviewers received a mix of documents selected for “contextual diversity.” This feature, integrated into Predict, allows the algorithm to keep finding and training against documents which are different from those already found through keyword search or seen by the reviewers in their initial rounds. You can read more about our unique contextual diversity algorithm on our website.

The review continued while the collection team added more documents. Since Predict can continually rank all the documents in the collection, there is no problem adding new documents during the review. As they are added, the documents are ranked and mixed into the total collection. To the extent they are similar to already ranked documents, they join the ranking in their proper place. To the extent they are different than what has already been collected, they become candidates for contextual diversity and can be included in the review sets for hands-on evaluation by the reviewers.

“Since Predict can continually rank all the documents in the collection, there is no problem adding new documents during the review.”
TAR 1.0 systems typically train against a reference set, which makes handling rolling collections difficult. To be representative, the reference set must be chosen randomly from the entire population and then carefully tagged by a subject matter expert. Then the training process begins with each round being measured by its effectiveness against the reference set.

If new documents are collected during the TAR 1.0 process, you have two options, with neither being ideal. Either you hope/assume that the new documents are similar to those already collected. Or you start again, discarding the initial reference set and its related training for a new round.

**Rolling Collections and Continuous Learning**

As mentioned earlier, through rolling collections over the course of several weeks, the Predict population grew to 3.6 million unique, rankable documents. As the review team found new types of responsive documents and learned more about the case, they could also use any other search and analytics tools available to keep searching. Every decision they made was continuously fed back into Predict to improve its ranking. When the review team ran out of relevant documents, they stopped the review and conducted a further systematic random sample of the entire population. Here is what they learned:

![Graph showing Predict Ranking and Linear / Random Review](image)
As you can see from the resulting yield curve, Predict was still pushing relevant documents to the top of the review pile, even after multiple rolling collections were added while the review was in progress.

Ultimately, the total review effort was about 500,000 documents, out of 3.6 million scheduled for review. Predict allowed the review team to achieve the requisite recall after reviewing only a small fraction of the population, which met the client's needs for both speed and efficiency. The team is now using Predict to help organize the review of all in-bound productions from other parties.

**Tokenizing Japanese Documents**

In the early days of technology assisted review (TAR), many questioned whether it was suitable for Japanese and other Asian-language documents. Indeed, for most TAR 1.0 engines, the answer was, and perhaps still is, a resounding “No.” After all, these products were designed to work on English-language documents that use spaces and punctuation to define word boundaries. For languages that do not follow Western syntax, the systems could not build the indexes required for them to work.

It is important to note that TAR systems don't actually understand the words they index and analyze. Rather they employ mathematical algorithms to determine the frequency and use of the words both in the documents and across the document population.

Japanese and other languages that do not use spaces between words often have to be “tokenized” (broken out into artificial words) before they can be indexed for search and analytics tools. Many earlier tools do this in a simple way, just taking two or three characters at a time. While this approach works okay for basic search, it can make analysis very difficult for TAR 1.0 systems.

TAR 2.0 systems such as Insight Predict employ special software to tokenize Japanese and similar languages a smarter way. They are able to analyze the text and break out actual words and word phrases, not just arbitrary groups of characters. Once the Japanese documents were properly tokenized, the TAR 2.0 process could index and analyze them more effectively.
Conclusion

This case presented a number of challenges. The collection was mostly in Japanese and contained a number of highly technical documents. The richness of the collection was low and it contained a lot of junk. The client was on a tight timeline for review but collections kept arriving on a rolling basis.

Despite these challenges, we were able to make use of the 10,000 documents the legal team had already reviewed to jumpstart the ranking process and accelerate the review. Even with the collection’s low richness, the team was able to find highly relevant documents many times faster than with any other approach. And because Predict never stopped learning from newly-reviewed documents, it continued to improve and help attorneys explore the collection even as new documents were constantly being added.

In the end, using Predict and its ability to support Continuous Active Learning, the client was able to cut the time and cost of its review by over 85%.

Postscript: A Supplemental Review

Because of continuing collections in the case, another 132,000 documents were loaded after we finished the final systematic random sample. But because we could use the existing judgments and Predict training, our client was able to get through that last batch of stragglers quickly.

The review of that late-arriving batch had to have a separate validation procedure performed to support the decision to stop the review when the review team no longer saw relevant documents. The procedure showed that the estimated recall on that last batch was 87%, even higher than the recall for the much larger, master collection that had accumulated over the rest of the review.