PRODUCT REVIEW

Reviewing Catalyst for Your eDiscovery Litigation Needs

By Brett Burney | May 15, 2015

The bad news: eDiscovery isn't going away.

The even worse news: eDiscovery projects will keep getting bigger because we live in a world of “big data” and digital packrats.

The solution: find a tool that can easily handle the ever-increasing data compilations collected in litigation matters and analyze everything quickly and efficiently. The document review tools developed decades ago, before we even used email, are woefully under-equipped to handle the data onslaught of today’s litigation matters.

Catalyst is one company that’s been on the forefront of the eDiscovery battles for years and they continue to push the boundaries of how technology can help lawyers face the struggling challenges involved in eDiscovery.

Getting Some Insight into Your Documents

Catalyst Insight was released about three years ago with an unorthodox twist—the company rejected the traditional “structured query language” (SQL) back-end found in most document review platforms and opted instead for an XML-based NoSQL platform. Put simply, this means Catalyst Insight can perform at a scale well beyond the limits of SQL platforms and immediately respond to any search request even against tens or hundreds of millions of documents.

Catalyst Insight is built for all things big: big data, big files, and big searches. You can cull down a collection of 20+ million documents to 140 search results in a matter of seconds. You’ll be hard-pressed to find another tool on the market that can accomplish that level of speed and fidelity.

Predictive Coding to Predictive Ranking

Catalyst has taken the same creative approach to providing tools for “technology assisted review” (TAR) with Insight Predict.
A typical TAR approach may re-use some older machine learning and text analysis algorithms to have a computer help identify the documents you need to review or produce. But the process takes time. A group of subject matter experts may have to look at a random sample set of documents (once all your documents are collected) and determine which ones are responsive to a litigation matter. The computer then identifies similar documents, tests the results, and repeats the process until the group is satisfied the computer is accurately identifying responsive documents.

But why wait for collections to finish and then wait for a group of subject matter experts to look through random samples of documents and repeat the process over and over? Instead, Insight Predict allows reviewers to start making judgments or coding decisions documents immediately on whatever documents are available, and the system continuously and constantly ranks and re-ranks the documents accordingly. That continuous analysis is also used to determine—on the fly—which documents the reviewer should look at next so that they find the good stuff as quickly as possible. This is what Catalyst (and others) refer to as “Continuous Active Learning” (CAL).

As John Tredennick, CEO of Catalyst wrote in his book, “TAR for Smart People,” the process can be compared to the way the Pandora Internet Radio works. Pandora has millions of songs in its archive but it has no idea what music you want to listen to. You start by giving Pandora the name of an artist or song that you do like. Pandora will then play a similar song and you give that song a thumbs-up or thumbs-down. Pandora uses your decisions to curate a selection of music that you really like, although occasionally it will still play something that you don’t like.
In a similar fashion, Insight Predict is constantly ranking thousands or millions of documents based on the decisions fed into the system. Even better, the process can continue even when new documents are introduced into the system (e.g. a rolling collection/production) without having to stop and start all over again.

**Predicting Successful eDiscovery**

To begin, documents are loaded into a Predict Project and you choose which decision fields (typically "responsiveness," but it can rank for any measure in the database) will be used for ranking all of the documents. The project administrator monitors the progress as reviewers declare documents relevant or non-relevant, and the Predict engine continuously ranks the whole population based on all the available review decisions.

Catalyst staff can help you read the graphs and statistics on the Insight Predict project dashboard so that you can see how the training is going, how the review is going, and make appropriate planning decisions. For example, one decision may be to generate a Yield Curve and establish a “cutoff” point as soon as the Predict engine is well-trained. In other words, with this approach you can establish when you’ve looked at a sufficient sample size of documents to confidently determine the responsiveness of the remaining documents. Once you are satisfied with the Yield Curve calculations, the administrator can assign document batches to a full review project where teams can code further or perform additional QC checks before production.
But there’s even another way with continuous active learning. With CAL, you can just keep reviewing in Predict and let it continue to improve the ranking it is using to decide what documents you should see next. To a reviewer, this just feels like a regular document review. But because Predict is continuously learning throughout the review, it allows reviewers to see all the responsive documents even faster than if they had stopped training earlier in the process. So for reviews where you want to have attorneys look at every document that gets produced, CAL can be faster and more efficient.

In any case, the system tracks the number of reviewed and un-reviewed documents, the precision and recall percentages, and then extrapolates the dollar cost of what’s been reviewed already versus what will be unnecessarily spent if you continue to review past any chosen cutoff point.

**Big Discovery for Big Data**

If you haven’t been involved in an e-discovery technology-assisted review yet, some of the information above may sound a bit intimidating and alien. But the stark reality is that TAR approaches to e-discovery are being embraced, and in some cases even expected when the document collection is impossible to search and review manually.

There are several TAR systems available on the market but Catalyst is forming the front lines of how the legal profession will be dealing with e-discovery in the coming years.

**About Brett Burney**

Brett Burney is Principal of Burney Consultants LLC where he focuses the bulk of his time bridging the chasm between the legal and technology frontiers of electronic discovery. Brett is very active in the legal technology community and was the Chair of the 2015 ABA TECHSHOW Planning Board. You can e-mail Brett at burney@burneyconsultants.com.