Digital Discovery & e-Evidence

BEST PRACTICES & EVOLVING LAW

http://ddee.pf.com

TALKING TECH

Automated Document Review Proves Its Reliability

By Anne Kershaw

Pushed by cost, time, regulatory and ethical considerations to embrace change sooner rather than later, law firms and clients are increasingly experiencing the impact of electronic discovery technologies. Staying ahead of the curve on these offerings is key to the effective management of discovery, providing the most reliable and cost effective case management for clients. While we have not yet reached the brave new world of completely automated document review, independent evidence suggests that automated techniques can do a significantly more accurate and faster job of reviewing large volumes of electronic data for relevance, and at lower cost, than can a team of contract attorneys and paralegals.

We discuss below a substantial study that we conducted, comparing automatic relevancy assessment to relevance assessments made by people. It demonstrated that using an electronic relevance assessment application and process reduced the chances of missing relevant documents by more than 90 percent.

Changing Landscape

Traditional methods of document review are typified by manual review using contract attorneys, entry level lawyers or paralegals. The individuals who are part of the review team are increasingly challenged by the sheer volume of data typically generated and stored by almost every organization that uses computer technology. Indeed, in some cases, it is simply not humanly possible to read all of the potentially relevant e-mail and documents within the time parameters set by the court.

Recent technological developments in the area of automated document review for relevance assessment are solving these problems, paving the way for profound and fundamental changes in the way discovery is conducted. In addition, technology can further level the playing field for smaller firms, by providing the ability to conduct large scale review with far fewer resources. While in the past lawyers may have been slow to embrace new technologies, all counsel would be well served to take early notice of the area of electronic document assessment.

Driving Change – Better Results for Less Cost, in Less Time

While law firms ultimately will derive many benefits from advanced document analysis technologies, large data producers such as universities, corporations, and government, are generally the leading proponents of their adoption. These data producers are driven in large measure by the enormous costs associated with conducting manual discovery in large document cases, which can easily encompass tens of millions of electronic documents.

Some companies are already starting to mandate that law firms use specific advanced technologies, even paying consultants to help make this transition successful; courts handling large cases may soon follow suit. In the future, more companies will, as a matter of course, tell counsel not only that they have to use technology, but identify which vendor they are required to retain. Law firms, large and small, would be well served to embrace these technologies before they are sent scrambling to do so by clients and courts.

Cost is certainly a principal driver in this shift. Automated document assessment solutions are cheaper, in most cases, than paying for an equivalent manual review capacity. Data collections often run into many gigabytes or even terabytes of data. Considering that one terabyte is generally estimated to contain 75 million pages, a one-terabyte case could amount to 18,750,000 documents, assuming an average of 4 pages per document. Further assuming that a lawyer or paralegal can review 50 documents per hour (a very fast review rate), it would take 375,000 hours to complete the review.

In other words, it would take more than 185 reviewers working 2,000 hours each per year to complete the review

within a year. Assuming each reviewer is paid $50 per hour (a bargain), the cost could be more than $18,750,000.

Electronic document review and assessment applications can now reliably identify the relevant documents first, and sort them according to subject matter. This dramatically reduces the volume of data requiring review by professionals for privilege and confidentiality and makes that review process substantially more efficient and cost effective.

"It is quite usual to see cases where we reduce the amount of data to be reviewed by 80 to 90 percent," reports Jonathan Nystrom, Vice President of Sales with Cataphora, a vendor of advanced electronic discovery services. "Only the time and cost savings possible using the latest electronic discovery tools make it even possible to undertake such a project," he adds.

A recent study that appeared in *Digital Discovery & e-Evidence* showed that, for a smaller case with 30 gigabytes of data, manual review could cost $3.3 million. The study described how a more advanced electronic approach could reduce that cost by 89 percent, to less than $360,000. (See "Document Analytics Allow Attorneys to be Attorneys," Chris Paskach and Vince Walden, *DDEE*, August 2005, page 10.)

Moreover, further cost savings can be realized by using the same technology, and often many of the same findings, across multiple cases over time. For example, a pharmaceutical company, for which litigation is a way of life, will often be required to produce very similar evidence in case after case. "Automation can measurably reduce these costs of doing business," comments Nicolas Economou, from electronic discovery services company H5 Technologies.

For large data producers, the consistent and repeatable processes provided by advanced review technologies are important, in addition to their accuracy, speed, and cost advantages.

Finally, the ability to see the fact pattern in the case earlier, thanks to the speed of automated review and the advent of electronic document analytics, provides better insight as to when early settlement might be appropriate, eliminating the costs of prolonging the matter unnecessarily. Such analysis also helps attorneys assess the benefits and trade-offs of producing documents in native format versus tiff images with fielded text.

Electronic Discovery's Old Guard

Many different types of tools have been developed over the years that provided limited support for electronic discovery. For example, a common approach has been to put imaged data (e.g. tiff files) and text into a database where the information can be examined using keyword searches.

Unfortunately, keyword searches are limited in their effectiveness. Not all documents of importance necessarily contain a candidate keyword and, at the same time, any chosen keyword will likely occur in many documents that are not of interest. As a result, documents of interest constitute a small minority of those located. The problem then remains, how to find the desired documents among the many that have been returned. Attempts to refine keyword searches by, for example, adding Boolean constraints (i.e., some combination of "ANDs" and "ORs"), do not usually provide much significant improvement.

The most advanced tools available today offer vastly improved capabilities. Legal teams can use such tools to locate relevant documents much more efficiently than ever before. And this evidence can be found much earlier in the proceedings. Getting more relevant information early in the process puts attorneys in a much better position to determine case strategy and gives them a much stronger basis from which to negotiate with the opposing side.

The State of the Art

Many vendors today provide the capability to use statistical techniques to determine which documents are "similar" according to specified criteria or exemplars and to group them together. This can help reviewers focus their efforts and provides huge time and cost savings over the course of a review. However, it is important to validate the accuracy of such automated categorization vis-à-vis the responsive specifications.

In many instances two documents may objectively be very similar to one another, yet one may be responsive and the other not. For example, in a particular matter, a document discussing the sale of a particular product may be responsive only if the sale in question occurred in the United States. Yet documents that relate to sales in the U.S. may be very similar to documents relating to sales abroad. In this example, it is very easy to see how two virtually identical documents, which would be grouped together by this technology, could fall on opposite sides of the responsiveness line.

Another approach is the use of what some vendors call "ontologies" or "word communities." They capture information about the words and phrases that model a particular area of knowledge. For example, in a case relating to alleged insurance fraud, an ontology might address particular industry practices that are potentially relevant to an investigation, or certain insurance-specific vocabulary that could be indicative of a responsive document.

Ontologies can provide a means of very accurately pinpointing relevant information. Equally valuably, they can be used to identify irrelevant materials, including junk e-mails, which can then be removed from consideration, thereby decreasing the amount of potential evidence that has to be reviewed. Additionally, much of the information captured by ontologies can be reused from matter to matter.

Contextual review is another example of advanced electronic document assessment. This technology uses the context between different documents to help reviewers deter-

mine the importance and relevance of a piece of potential evidence.

"Traditionally, context has meant looking at context *within* a document," comments Cataphora's Nystrom. "By contrast, we now have the ability to look at context in the form of the relationships *among* documents. Seeing potential evidence in the context in which it was originally created and used makes it much easier for reviewers to make accurate assessments of its relevance and importance, and to do so very quickly."

Some of these tools also provide litigation support managers with increased control over the review. They can then ensure that the review is completed on time and within budget. To help managers do this, advanced tools can provide information about how much of the evidence has been reviewed, and how much remains. Review managers can then determine whether they have enough resources to get the job done on time and to make adjustments at the earliest possible opportunity. It is even possible to monitor the speed and effectiveness of individual reviewers, tracking how much evidence each reviewer has processed. Review managers can also see which reviewers are finding the largest numbers of relevant documents, and how accurate their review decisions are.

Electronic Document Assessment for Relevancy Really Works

Historically, human review has been the gold standard for initial relevancy assessment. Yet it was rarely, if ever, tested for accuracy. The advent of electronic relevancy assessment processes and applications now allows for the comparison of these techniques against human review. We conducted such a study and found not only that the electronic assessment for relevancy was highly accurate, but also that people reading documents to assess relevancy missed close to half of the relevant documents.

Our study began with a set of 48,000 documents, which were to be coded for relevance to three responsive categories. The software was set up in accordance with the vendor's standard practices, which included interviewing the attorneys and reviewing documents to gain an understanding of the relevance criteria for the case and training the software accordingly. In parallel, six reviewers were trained to conduct a manual review of a stratified random sample of 43 percent of the corpus.

The software and the reviewers separately reviewed the documents and the results were compared. We assumed that where the software and the humans agreed, the determination was correct. Where there was a discrepancy (a document marked responsive by one approach and not by the other), the document was re-examined by the same reviewers to determine (in some cases with some debate and arbitration) who was correct, the software or the human reviewer.

At the end of day, after all the numbers were crunched, the human reviewers were shocked at how many documents they missed and were similarly startled at how well the software achieved the objective of locating relevant documents. Across all three codes, the software, on average, identified more than 95 percent of the relevant documents, with a high of 98.8 percent for one of the codes. The people, on the other hand, averaged 51.1 percent of the relevant documents, falling as low as 43 percent for one of the codes.

These findings makes sense considering that document review work is extremely difficult, that people have subjective views of relevancy, and people can be easily distracted from the work by fatigue or thoughts of lunch and other matters. The software process, on the hand, consistently assesses every document and never gets tired.

In sum, the results of our study demonstrated that the use of a particular software application and process reduced the risk of missing a responsive document by 90 percent. Moreover, the effectiveness of the electronic process improves as it is tweaked throughout the quality assurance process. These results may be surprising to those who have an abiding belief in the quality of traditional manual review, but they are probably an accurate — maybe even optimistic — reflection of the performance of an average review room, particularly if the case is large and complex and review is being conducted, as it so often is, against an aggressive deadline.

The legal world may not yet be ready for fully automated review, and there will long remain a role for expert human review. Nevertheless, advanced technologies can be used to focus review efforts on those documents that are most likely to contain relevant information. At the very least, such tools can be used with some confidence to root out obviously non-responsive materials, allowing review to focus on what is left. That alone can provide considerably increased efficiency, reduced costs and superior results.

What this Shift Means for Lawyers

The newest technologies open the door to successful handling of much larger volumes of electronic evidence than has ever been possible before. Faced with the advent of these tools, attorneys have the choice to either embrace them, or take the risk that competing firms will take business away from them.

Automated document review and analysis provides significant new opportunities for attorneys in law firms and in corporate legal departments. Legal review can be a more efficient, less costly, and a more proactive process that aids the legal team in managing the case.

There is every sign that the competition will become more intense. Technology can level the playing field by giving smaller firms the same review capability as larger firms, and business as usual will not be an adequate response. All law firms, large and small, must prepare for the impact of the new technologies.

Anne Kershaw is the founder of **A. Kershaw, P.C. // Attorneys & Consultants,** a nationally recognized litigation management consulting firm providing independent analysis and innovative recommendations for the management of all aspects of volume litigation challenges.

Ms. Kershaw provided electronic discovery survey data and testimony before the Federal Civil Rules Advisory Committee. In addition, she is a principal author of *Navigating the Vendor Proposal Process: Best Practices for the Selection of Electronic Discovery Vendors* and a contributing editor to *The Sedona Conference Glossary For E-Discovery and Digital Information Management (May 2005 Version)*, both projects of the Sedona Conference® Working Group on Best Practices for Electronic Document Retention and Production RFP+ Group (www.thesedonaconference.org).

Further information regarding Ms. Kershaw's resume, career and practice is available on www.AKershaw.com.