Let’s Get Relevant

Using Document Analytics To Reduce Total Discovery Cost

By Mike Kinnaman

Tremendous volumes and increasing varieties of electronic information create onerous burdens for corporations dealing with discovery requests, internal investigations and response to regulatory agencies.

To help combat this technology burden, corporations are employing document-analysis technology to accelerate the identification of relevant information. An emerging best practice, this approach yields considerable cost and time benefits that help law firms reduce discovery risk and expense for their clients.

Until very recently, the most common approach to document discovery and review was to convert all documents to a text-searchable image format such as .TIFF or .PDF. This enabled keyword searching across the collection to identify potentially relevant material. Once the relevant material was identified, it could easily be redacted and Bates numbered. As the amount of discoverable electronic information increased, however, this approach quickly became extremely cost-prohibitive for corporations because the per-page fee applied to relevant and irrelevant documents. When up to 95% of the documents are likely not to be found relevant to the discovery request or investigation, it is incumbent on the law firms to identify tools and techniques that can limit the expense of irrelevant material for their clients.

“The cost of e-discovery poses an economic threat to any company facing litigation,” Stephanie Mendelsohn, partner at Reed Smith, says. “Our firm increasingly looks for ways to help clients manage the costs and risks of e-discovery. In cases with large volumes of potentially relevant documents, document analytics can be an effective and strategic tool for managing and meeting electronic-discovery obligations.”

THERE MUST BE 50 WAYS TO REVIEW 1 MILLION DOCUMENTS

Much of electronic document discovery’s burden can be recognized in its complexity: Each matter encompasses multiple custodians with a myriad of combinations of file types and data sources. This complexity creates painful chain of custody issues and challenges for even the most sophisticated project manager. While most IT managers scoff at the difficulty of managing 100 GB of data (“100 GB? Big deal! I’ve got 100 GB on my new iPod!”), the 100 GB discovery request will result in the time- and resource-consuming attorney review of hundreds of thousands of documents.

To help winnow down such large populations of electronic information, e-discovery vendors and consultants offer a number of programmatic culling tools and techniques such as:

• Keyword searching;
• Culling by metadata (eg, date saved or sent);
• Eliminating exact and near duplicates; and
• Concept searching.

Yet even after using the most sophisticated programmatic culling techniques, the vast majority of the collection will be irrelevant to the investigation or discovery request.

Context is what makes programmatic identification of relevant material unsuccessful. While a
keyword search can identify thousands of matches to a word, the meaning of that word can be wildly varied depending on its context. A simple way to understand this is to Google the word “diamond.” You’ll discover about 40 million different search results related to jewelry, baseball, card games and shapes — in addition to hundreds of product and company names.

“The challenge in discovery is that whatever you are looking for — key facts, hot documents, privileged materials — is often buried in massive volumes of irrelevant material that consumes attorney review time and drives up costs,” says Kathy McFarland, of counsel at Lovells. “The more efficiently you can isolate what is important from what is not, the more effective you can be for your client.”

**DOCUMENT ANALYTICS AND ELECTRONIC DISCOVERY**

While much of the vendor lore in the e-discovery space consists of scary “smoking gun” tales, the more frequent and costly challenge of e-discovery and document review for corporations is the tedious process of slogging through endless sets of e-mails, spreadsheets and Microsoft Word files. Regardless of approach, the initial objective with each and every document is to determine its relevance to the matter. The more efficiently and cost-effectively a reviewer can make that initial determination, the more time and money will be preserved for case strategy and negotiation.

An emerging best practice in the management and review of large electronic-document collections — particularly under compressed response timeframes — is the use of document analytics to help attorneys quickly identify relevant material in its native format. At first glance, this approach might seem more complex because it introduces a new step into the review process. Law firms are quickly finding, however, that this extra step is a shortcut that preserves valuable time for strategy while reducing the overall expense for their clients. Once identified, the relevant documents can be transferred to a secure online repository for further legal review and case management. If required, the documents can at that point be converted to a common format such as .TIFF or .PDF. This immediately saves the client money by avoiding the unnecessary conversion fees associated with irrelevant materials.

To understand the role of document analytics in the legal review process, it is helpful to define document analysis, document analysts and document analytics.

Document analysis is the process of understanding what a document “is” in relation to a higher-order context. Document analysis often involves such tasks as determining a document’s authenticity, author and creation date. In the legal arena, it also implies identifying the document’s significance or relevance to the investigation or the discovery request.

Document analysts are professionals — such as attorneys, paralegals and fraud examiners — accomplished in the art of document analysis. Fraudulent activity that is transacted across corporate e-mail servers is often not immediately identifiable in an investigation because of code words or acronyms designed to fly under the supervisory radar. These situations require an experienced document analyst to identify such

language and transactions.  

Document analytics is the emerging technology field devoted to helping document analysts with the challenge of document analysis. In the legal arena, document analytics combines familiar e-discovery techniques with newer analytic capabilities in a single software application to help automate and simplify common electronic-discovery tasks. The result is a complete and integrated toolset that provides the document review team with multiple approaches to understanding the document collection and identifying relevant material.

An example of a document analytics method is the visualization technique known as “document mapping.” Document mapping is accomplished by extracting the concepts (nouns or noun phrases, e.g., people, places and things) that compose each document and applying them to better understand the varied relationships between all of the documents in a collection. These relationships are visually plotted on a map according to each document’s conceptual similarity to others. In other words, a relevant document is plotted on the map in close proximity to documents that share similar concepts. Rather than leafing through a large collection of documents page by page (or screen by screen), this approach enables the document analyst to see the conceptual relationships at play in the collection and develop a more efficient review strategy.

Advanced applications of document mapping go even further by permitting the document analyst to filter her view of the collection through emphasis or suppression of concepts — or both — by selection or example. In the selection model,
the document analyst can re-plot the document map by emphasizing important concepts or suppressing less meaningful ones. In the example model, the document analyst can reorganize the collection around a relevant document to quickly see what documents are most conceptually similar to it. In both examples, the document analyst can quickly distinguish relevant from irrelevant material by dynamically generating a more meaningful concept map based on her understanding of relevancy in the matter.

“By clustering related materials into a document map, attorneys set aside irrelevant documents while tagging groups of highly relevant documents fast,” McFarland says. “Lovells uses this ‘eyes on’ advantage to tightly focus on selecting key documents while distinguishing relevant from irrelevant materials.”

GET RELEVANT: EVALUATE TECH IMPACT ON TOTAL DISCOVERY COST

Despite the wonders of modern technology, there is still no pill-form panacea that adequately addresses the sum of electronic discovery’s intricacies and challenges. Electronic discovery — whether in response to a discovery request, in an internal investigation or just for weekend fun — is an arduous process that requires capable professionals to guarantee a successful outcome. What technology can do is help the professionals fight technology with technology by increasing their productivity and streamlining the e-discovery process. The desired outcome of any e-discovery technology investment should be a reduced discovery bill and the more expeditious completion of discovery.

A simple method for assessing the technology impact on the total cost of discovery is to break down the overall expense into two categories: technology cost and attorney review cost. The attorney review cost can be determined by multiplying a blended hourly attorney rate by the number of total hours required to review all material. The technology cost is the aggregate of fees associated with collecting, processing and hosting data in addition to any software expenses such as online review, case management or document analytics — or a combination of these. Any project management or consulting fees related to this technology should be included in the technology cost.

One of the advantages of using technology to fight technology is that valuable details related to electronic-document processing and review can be tracked and measured. As a result, law firms can monitor important metrics to better manage client expectations and evaluate the worth of different approaches. A few key metrics that can help corporations and their law firms better understand the impact of a technology investment are:

- **Attorney document decisions per hour.** How quickly and accurately can the reviewing attorneys determine a document’s relevance? As the initial objective with each document is to determine its relevance, the rate at which this can be determined can generate significant cost-efficiencies for the client.

- **Cost per non-relevant document.** Reducing expenses related to irrelevant material presents one of the largest opportunities for client savings. How much does the client spend on each non-relevant document, including the technology cost and the attorney review cost?

While technology can be a powerful asset in mitigating electronic-discovery risk and expense, it is important that corporations and their law firms truly understand the impact of their technology investments. e-Discovery success starts by understanding the total cost of discovery — the cost of attorney review time plus the cost of technology. The total cost of discovery is an essential foundation for creating powerful metrics, such as cost per non-relevant document, to illustrate how technology reduces the discovery expense and preserves valuable time for case strategy.

The use of document analytics software early in the e-document review process can help the review team more quickly identify relevant information while minimizing unnecessary expenses related to the review and management of irrelevant information.

“In a recent matter, the use of document analytics helped us quickly isolate the most important items and cut the overall time it took to review the documents by one half,” McFarland says. “Based on our analysis in this matter, I believe the technology resulted in a 75% cost-savings for our client.”