Relevance Ranking is the Key to Early Case Assessment

By Warwick Sharp

December 1, 2006 will go down as a watershed in not one, but two high-profile professions. On that date, both the U.S. legal community and the enterprise IT community were fundamentally changed as the federal government published its amendments to the Federal Rules of Civil Procedure, codifying the rules for e-discovery as applied in U.S. civil litigation.

Five years later, e-discovery has turned into a billion-dollar industry, and it now represents perhaps the biggest procedural challenge facing litigants and their legal representatives. E-discovery costs continue to grow rapidly, fueled by the ballooning volume of digital content of all kinds generated in the course of doing business.

Corporate America has been heavily impacted by e-discovery mandates. Many companies, especially those in notably litigious industries, have been compelled to dramatically increase their volume of stored documents. Consequently, the steps in e-discovery - pre-processing, processing, analysis, review and production – have become hugely expensive, and the costs continue to go up.

To help stem this tide, a practice known as Early Case Assessment (ECA) has emerged. Simply put, ECA is the process of analyzing a subset of case-related information in order to determine whether to pursue the case.

ECA attempts to assess the potential risks and costs of the case, taking into account the strengths and weaknesses of the corporation's position and the potential scope of litigation review that would be involved. With this assessment in hand, the corporation can decide whether or not to continue down the litigation track. Through an early decision to settle, the corporation can avoid the hugely expensive costs of e-discovery and document review that litigation involves. According to some estimates, 70 percent of a total litigation budget can be consumed in e-discovery.

ECA can not only help determine exposure relative to a matter, but it can facilitate setting an appropriate e-discovery budget should the client choose to pursue the lawsuit. Thus, properly done, ECA can either avert or justify millions of dollars in legal costs.

ECA CHALLENGES

A whole new class of service providers and solution vendors has sprung up, all claiming to “own” the new ECA space, and numerous theories and approaches have been advanced about how ECA should be performed. All agree that technology is key. The question is, which technology is most effective?
Traditionally, legal teams have been taught that an exhaustive reading of all relevant documents is mandatory, but obviously in the age of multi-petabyte data stores that's no longer feasible. The central challenge in ECA, therefore, is to identify the highest value materials as expeditiously as possible, so the most informed decisions can be made.

Opinions differ on the best methodology for doing this. Because communication now takes so many forms and the most revealing documents are not necessarily obvious, a wide range of search, clustering, classification and analytical solutions have been proposed.

Keyword matching has been the default strategy. As applied to ECA, keyword matching begins with a list of terms compiled by the lead attorneys. A search is performed once the broad categorization of documents has been identified based on date ranges, data sources, etc., and system files, duplicates and other superfluous documents have been removed.

Such a search is only as good as its keywords, and somewhat paradoxically keywords tend to be both over and under-inclusive. Since keywords ignore context, keyword searches often flood an ECA team with irrelevant documents. On the other hand, studies have shown that keywords regularly miss up to 80 percent of documents relevant to a matter.

No less important, keyword searches don't prioritize results. Documents are either "in" or "out." For legal teams that are constrained by deadlines or manpower, and for which a large body of material remains after sifting, the lack of prioritization can render their ECA exercise ineffective.

Keyword matching may be the most widely-used strategy in the first wave of ECA, but it is by no means the only one. Some strategies are advanced as primary search techniques in their own right, while others are intended to be applied in combination with more conventional strategies to isolate vital, even "smoking gun" documents.

Sentiment detection and analysis looks for instances in which emotional words enter routine business communications — the thinking being that matters of potential culpability typically generate more spirited and intense exchanges.

Social network analysis looks at who and how people are relating to each other. The technique flags communiques that are outside the realm of normal workday interaction — for example, a salesperson who is repeatedly emailing a security staff member.

Such techniques, while helpful, still may not root out the most essential
elements in a case. The information that legal teams are looking for is not necessarily in communications where sentiments are expressed, and it does in fact often reside in exchanges between people who communicate often.

There are, however, new ECA technologies that seek to leverage the most astute and intuitive resource available: the skilled and informed attorney.

A NEW APPROACH
Relevance ranking has emerged as a way to make sense of the huge volumes of content encountered during ECA. As its name indicates, this process evaluates and prioritizes documents according to their degree of relevance to the case. Because relevance ranking grades documents, placing the most pertinent documents first, it allows counsel to review the most applicable documents in rank order, a key prerequisite of effective ECA.

A major advantage of relevance ranking is that it amplifies and automates the judgments of lead attorneys, rather than attempting to replace them. The technology “learns” to parse documents in an iterative process that feeds off the lawyer. By continuously interacting with the technology during the learning phase, attorneys can be assured that their critical knowledge is leveraged accurately and to the fullest extent possible.

Furthermore, the fidelity of relevance ranking to a designated legal team’s collective judgment provides valuable justification in court, should future challenges to the team’s ECA process arise.

In relevance ranking, specialized software identifies a test set of perhaps 50 documents that an attorney close to the case reviews and codes according to relevance. This process is repeated multiple times, each time with a different set of test documents. With each repetition the software learns more from the expert and its accuracy score is improved, until a threshold of competence is achieved.

This process may take a couple of days. At that point the application is turned loose to perform its automated evaluations on the entire document set. In a short period of time—and at a rate of several million documents per day—the collection is thoroughly evaluated and ranked. Formal relevance scores are calculated for each document, with the highest rankings given to those files that must be reviewed first.

Relevance ranking allows practitioners of ECA to make the best use of available human assets, since senior attorneys can examine the most important documents while associates tackle the mid-ranked pile. Those materials with the lowest scores are set aside, although they remain available should changes in the proceedings require a re-evaluation. This document hierarchy is one of the key drivers of value for attorneys who use relevance ranking.

Despite the advantages of relevance ranking, some attorneys still hold to the notion that ECA sorting technology can never replace the combination of keyword matching and formal human review. To test this belief, two formal studies have been undertaken, one using the 2008 TREC (Text Retrieval Conference) data and another by Equivio and Epiq Systems. In both cases, results proved the efficacy of relevance ranking.

In the Equivio/Epiq study, automated relevance sorting was a 91.4 percent match to a document set produced by a human review team. Moreover, an independent human “oracle,” retained to judge a sampling of documents in which the automated vs. manual approaches differed, found that the automated alternative was correct in 77 percent of the cases.

IMPLEMENTATION ALTERNATIVES
While no technology can be the complete and unique solution, relevance ranking is quickly gaining traction in the ECA arena. For many, the question is not whether to adopt relevance ranking technology (or any other ECA analytical software), but rather whether to buy or subscribe to it.

It’s important to remember that good ECA technology is complex. If you can justify the cost within your organization, retaining on-premise software can create maximum ROI. Gartner, Inc., a leading information technology research and advisory company, recommends that IT departments in companies with five or more $1 million-plus cases per year should budget at least $500,000 for software services and one full-time equivalent employee. Such users have reported a very short payback period—in the order of three to six months—for e-discovery software investments.

For those with more restricted budgets, the on-demand Software-as-a-Service (SaaS) model may be a better solution. The real question, once monetary issues have been resolved, is likely to be where your e-discovery budget should be invested. ECA technology has been shown to provide excellent return because it can prevent a client or legal team from going down the track of expensive litigation in cases where cost or risk are not justified. As a strategic investment, ECA technology is clearly near or at the top of the list.

PROPORTIONATE AND EFFECTIVE
As ECA continues to mature as a practice for managing the complexity of e-discovery, both its definitions and its strategies will evolve. Whatever protocols emerge as best practices, relevance ranking is a technology that both IT departments and legal teams ought to consider. Its ability to extend the intellectual reach of senior attorneys, combined with its demonstrable advantages as a proportionate and effective application of resources, makes it a preferred choice for the needs of Early Case Assessment.

Warwick Sharp is a co-founder and vice president of marketing and business development at Equivio, a provider of analytical solutions for e-discovery.

Warwick.Sharp@equivio.com