Published by Pike & Fischer, a BNA Company

Digital Discovery & e-Evidence

BEST PRACTICES & EVOLVING LAW

http://ddee.pf.com

Reprinted from Vol. 7, No. 1 | January 2007

VENDORS & TECHNOLOGY

When Less Becomes More: Making the Case for a Discovery Strategy

By Anne Kershaw

Somewhere in a database of tens of millions of documents is the "heart of relevance." This heart is only pierced with the right strategy, the right search engine, and the right people. This glorious heart will likely be less than 50,000 documents and would have been produced with appropriate privilege and confidentiality review in just a few weeks. We would know who the key players really are and what keywords really define the corpus. Relevance would be consistently applied with an inherent understanding of what the document universe has to offer. In the end, litigators would have confidence in the process and will know the essence of the documents well before the first "meet and confer."

With the *right people, strategy, and tools*, one can obtain a cost-effective process to determine relevancy and expand these findings "outward." This is in contrast to the old school, expensive, and burdensome process of brut force manual document relevancy and coding review. Like fighting fire with fire, technology, such as electronic mail systems producing millions of messages have overburdened the litigation process and it will be technology that will offer us the solution, provided we use an appropriate, reasoned strategy.

The federal rules governing electronic discovery, combined with several decisions following *Zubulake* and its progeny, underscore the need for attorneys and their clients to gain a greater understanding of their data and electronic discovery processes very early in the case, to avoid pitfalls along the way. Early, in-depth data analysis leads the way to meaningful meet-and-confer and Rule 16 scheduling conferences.[1] This should be viewed as a significant opportunity for large data producers because an agreed-upon or court ordered discovery plan is the best protection against problems such as unexpected evidentiary issues, production of privileged documents, added expenses, and court-imposed sanctions.[2] To get an agreement or order, lawyers will need to prepare targeted and defensible discovery plans and prove that their proposal is the best first shot at getting relevant evidence. *It is now clear is that attorneys will be forced to decide whether automated document review should be part of their discovery plan.*

Success in the execution of a strategy and technology driven discovery plan will be defined first as being able to sell this plan to your adversary, and if not successful there, to your judge. You will need to prove at the outset that not only did you use a reasonable process, but that you used a smart, thoughtful, and technologically reliable process, to find the data and documents that matter the most first. If you start with the heart of relevance, any additional requests will subsequently be focused and limited.

Importantly, the credibility of your representations will be constantly maintained. This not to suggest that anything should be "thrown away"; the goal is to get the cream to the top and produce it first.

Clearly, to convince everyone else, you need to be open and transparent about your search strategy and methodology, and to do that, you need to understand not only what you are looking for and why, but also the effectiveness of the searching tool that takes you there. Some level of technical understanding will be required when less becomes more.

The Right Strategy and the Right People

Strategy is often defined as a long-term plan of action designed to achieve a particular goal. In this case we are talking about developing a strategy for how best to assess and use search technologies against a whole dataset, to get to the heart of the relevant material in that data set, so that discovery starts with a manageable and informative document population. The development of a document search

and analysis strategy that works requires a complete understanding of the litigation and trial process, an appreciation for the organization that created the data (both its culture and language), and direct input from lawyers responsible for the case.

Part of the planning process is to invest time reading the pleadings and work product, deciphering organization charts, and understanding the issues of proof contained within the likely jury charges. Obviously, any documents already collected and considered relevant are important as exemplars of what may be in the larger document collection.

This plan will require a thorough understanding of the law, claims, timing considerations, subject matter, and the client. Documenting this learning and investigation stage is very important, since your adversary and judge will need to comprehend why certain search paths were eliminated in view of claims and other case constraints, while other paths were intensely mined due to their relevancy.

The Right Search Tool

Software advances now provide sophisticated document analysis capabilities well beyond "key word" searching, and software can now evaluate and determine relevance. Documents can also be "clustered" based upon an intrinsic understanding of the similarity of documents. These advancements profoundly change discovery in a way similar to the impact that on-line research had on the practice of law.

While there is a clear role for manual and intensive review of "hot documents," attorneys who embrace this new software paradigm will gain a unique competitive advantage in addressing their adversaries and the court. The shift will simply be that before initiating attorney review, the document corpus will be subjected to a rigorous and intelligent search and data analysis — either by the software vendor's consultants or experts on your team — to eliminate irrelevant documents, find the most relevant documents and to determine potential keywords. Once this much smaller set is defined, traditional review can commence, yet even here a remarkable shift will be noted. Now that the remaining relevant material is organized by some type of topical "clusters," reviewing lawyers quickly become subject matter experts, increasing review speeds and making these attorneys valuable assets to the case.

How do you verify and test the effectiveness of various technologies and solutions, and determine which vendors, if any, are most appropriate for a case? There is considerable confusion in the fast-evolving vendor playing field as to what various methods and tools are capable of doing and how well they perform. Further complicating matters is that lawyers should expect judges to have little tolerance for unintentional under- or over-production during discovery due to claims of vendor error or technological processes gone awry. Even in this world of less is more, litigants will need to exercise caution in choosing a technology solution and they should seek guidance and recommendations from trustworthy industry initiatives, such as those offered by the Sedona Conference.[3]

Perhaps the best — if not the most critical — component of the vendor selection process is benchmarking. Benchmarks that utilize independently verifiable methods and academically supported data can translate into evidence that will not only help you determine which vendor is best suited for your needs, but will also support a case for legal defensibility.

Three Cornerstones of Document Review Benchmarks

A proper benchmark for any high-volume review project will incorporate three variables: cost, time to completion, and accuracy. All three factors are variables in all review projects and often represent competing priorities. Benchmarking will help determine the optimal equilibrium among these variables.

Cost: In most cases, using technology to cut to the core of relevance, or using automated document assessment solutions, is much less expensive than paying for an equivalent manual review capacity. Search-driven review processes also have the potential to provide huge economies of scale, especially for companies involved in repetitive litigation. A cost benchmark should incorporate all discovery costs from the first document collected to the last document reviewed.

Time: Reducing the extent of the document review to that which is relevant unquestionably reduces review costs. In addition, the ability to see the fact pattern in the case earlier, thanks to in-depth searching by the right people 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.

Accuracy[4]: This is the most critical component in a benchmarking exercise, and deserves far greater discussion than is within this article's scope. Most organizations are focused on cost and time factors, perhaps because independent evidence on accuracy of e-discovery tools in identifying relevant information is so sparse. Indeed, it is not practicable to conduct independent tests that benchmark accuracy in any given matter. However, it is crucial that businesses *consider* accuracy in the benchmarking equation.

First of all, the level of accuracy obtained from a particular document review method directly effects variables of both cost and time (i.e., analyzing irrelevant documents creates substantial unnecessary costs and significantly delays time to completion). Moreover, it directly affects exposure to risk, because low accuracy increases the probability of failing to retain, produce, or find the relevant data.

Benchmarking Vendor Performance in the Document Review Context: Case Studies

When faced with a discovery request, lawyers need to know what information they have and to determine what information is relevant. Inevitably, they will encounter the

question of whether, or to what extent, automated review and assessment technologies and methods are accurate and effective.

Recently, independent studies have demonstrated that technologies can effectively automate the document review process. The first study of this kind demonstrated the impact of automated review and assessment technologies on cost, time, and accuracy. The results clearly pointed to these technologies as being more accurate and effective than manual review.

This study[5] evaluated an automated relevancy assessment and review process (offered by H5 Technologies) for a Fortune 50 company and a very large data producer. The results of the study demonstrated not only that the automated process was highly accurate (finding close to 98 percent of the relevant documents), but that the human review team missed close to half of the relevant documents. Crunching the numbers demonstrated that the automated relevancy assessment reduced the chances of missing a relevant document by 90 percent compared to human review.

The update to this study, to be produced in the spring 2007, will compare several vendor document review processes with a manual process conducted to respond to a Department of Justice "Second Request" subpoena. Again, we will be measuring the cost, time, and results of electronic document review processes against the results of manual review. These studies are being conducted by The E-Discovery Institute, an independent entity with a prestigious advisory board composed of experts in artificial intelligence, judges, as well as well-known and skeptical plaintiff and defense lawyers. The results of the study will be published and distributed to the public at no charge.

Call to Action

The law requires all sides to a dispute to be reasonable in their efforts to identify documents as potentially relevant. What is "reasonable discovery," however, is being challenged and redefined in the age of electronic discovery.

Lawyers and judges must understand the effectiveness of search strategies and the capabilities of automated relevance assessment systems, clustering methodologies, and the differences among them in order to determine the appropriate applications, if any, to their case. The scale of electronic discovery productions will make this an absolute must. Equally important, they need to understand how to use these technologies to ascertain the "heart of relevance" — and explain how they got there — to convince their adversary and if necessary, the judge, that less data truly is more, because it's the right stuff, not just a lot of stuff.

Anne Kershaw, Esq. *is an experienced litigation attorney and the founder of A. Kershaw, P.C. // Attorneys & Consultants (www.akershaw.com), 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 (May2005 Version), both projects of the Sedona Conference® Working Group on Best Practices for Electronic Document Retention and Production RFP+ Group. She is also co-founder of the eDiscovery Institute, a not-for-profit e-discovery research organization operating under an advisory panel consisting of prominent academics, judges, and electronic data discovery experts, and a faculty member of Columbia University's Master of Science program in Technology Management. Ms. Kershaw is also on is an Advisory Board Member of the Georgetown University Law Center's Advanced E-Discovery Institute.*

Special thanks go to Samuel H. Solomon, CEO of DOAR Litigation Consulting (sam@doar.com) for his insights and assistance in the development of this article.

(Endnotes)

1. Under the new rules, parties are now required to address issues relating to ESI early in the proceedings. According to the Committee Notes, Rule 16(b) was amended to alert the court to the possible need to address special handling of issues relating to the discovery of in the pre-trial conference stage. As amended, Rules 16 and 26 require that the parties meet and confer early on to discuss discovery issues (including the disclosure, preservation, and form of production of ESI) and to develop a discovery plan.

2. Numerous decisions establish that lawyers are not immune from sanctions for mistakes that lawyers attribute to their e-discovery vendors. *See, e.g., Coleman (Parent) Holdings v. Morgan Stanley*, 2005 WL 679071 (Fla.Cir.Ct. March 1, 2005); *Phoenix Four v. Strategic Resources Corp.*, 2006 WL 1409413 (SDNY Mary 23, 2006); *Residential Funding Corp. v. DeGeorge Fin. Corp.*, 306 F.3d 99, 108 (2d Cir. 2002); *In re Worldcom*, No.CIV.02-3288, 2004 WL 768573 (SDNY April 12, 2004). Note, however, that courts have generally limited imposing attorney sanctions to cases involving gross negligence, willful misconduct, or intentional destruction of evidence on the attorney's part.

3. *See* The Sedona Conference, "Best Practices for the Selection of Electronic Discovery Vendors: Navigating the Vendor Proposal Process" (July 2005); www.thesedonaconference.org.

4. Since we're on the topic of accuracy, let's define our terms. Accuracy is not a meaningful Information Retrieval term of art. The two most important measures used to gauge the overall effectiveness of Information

Retrieval systems (such as document review systems) are ***precision*** and ***recall***. Precision is a measure of how many of the retrieved documents are actually relevant, and recall is a measure of how many of the relevant documents were actually retrieved. A basic understanding of the measures of precision and recall is essential to understanding electronic discovery processes. It is also a practical necessity. For example, you can't do a cost benchmark without understanding precision and recall because precision and recall affect all downstream costs. (For a fuller discussion of these terms, see the preceding article, "Searching in All the Wrong Places," by Bruce Hedin, Ph.D., at pp. 9-11.)

5. For a summary of this study, *see* Kershaw, "Automated Document Review Proves Its Reliability," *Digital Discovery & e-Evidence* (November 2005). Although this study was conducted nearly two years ago, to date there are no other independent published studies that compare an automated approach to document review with conventional manual review.