E-Discovery Search Goes Under the Microscope

William Uppington

Over the last two years, the practice of electronic document search during discovery has received increasing scrutiny in the courts. In particular, the use of e-discovery search technology has been central to the outcomes of a number cases including *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 2008 WL 2221841 (D. Md. May 29, 2008), *Perfect Barrier, LLC v. Woodsmart Solutions, Inc.*, 2008 WL 2230192 (N.D. Ind. May 27, 2008), and *In Re Seroquel Products Liability Litigation*, 244 F.R.D. 650 (M.D. Fla. Aug. 21, 2007), to name just a few. In each of these cases, ill-advised, but common-place search strategies contributed to adverse outcomes. In Victor Stanley, an under-inclusive keyword search for privileged documents led to the waiver of attorney-client/work product privileges. In *Perfect Barrier*, an over-inclusive search for responsive documents led, in an ironic twist of fate, to increased e-discovery costs for the plaintiff who had initially proposed the search terms. In *In Re Seroquel*, the judge's disapproval of a defendant's search methodology contributed to a finding of sanctions against that defendant.

That e-discovery search is increasingly featured in discovery disputes should not come as a surprise. The increased scrutiny is a direct result of the massive increase in the amount of electronically stored information (ESI) that is affecting all aspects of e-discovery. As the amount of ESI grows, search technology represents one of the few mechanisms that can be used to stem the massive increase in electronic discovery costs that would inevitably occur without it. Since the amount of ESI looks set to continue increasing, the need to use search technology will only become greater.

The combination of an increasing need to use search technology within e-discovery and the increased scrutiny of common search practices puts today's e-discovery practitioner in a challenging position. How can one conduct a search within discovery in a way that helps manage the costs of e-discovery but is defensible? This article will discuss the challenges with today's search technologies and argues that one way to both reduce costs and improve defensibility is through increased transparency in the search. To this end, this article will introduce the concept of "Transparent Search" and explain how greater transparency can help litigators be more prepared for the scrutiny of a search that is yet to come.

Challenges in E-Discovery Search

The basic challenge with all search technologies including the most commonly used today, a keyword search, is that searches tend to miss relevant documents (i.e., be under-inclusive), and find non-relevant documents (i.e., be over-inclusive). In other words, a search is not perfect. And it is these imperfections that contribute to the disputes over the practice of searches in e-discovery. So the key question is how to conduct searches that are defensible given these imperfections? Fortunately, Judge Grimm provides some guidance on this.

Collaboration and Defensible Best Practices in E-Discovery Search

In *Victor Stanley*, Judge Grimm suggests the way to defensibly address the challenges and imperfections inherent in e-discovery searches is for the searching party to undertake one of the following two approaches:

- ***Collaborative Search Approach:***

  With this approach, the producing party would "confer with the opposing party in an effort to identify a mutually agreeable search and retrieval method." If both sides agree on the search methodology, then this largely removes the defensibility question as there is no party to question the methodology.

- ***Defensible Best Practices Approach:***

  Judge Grimm indicates that in the absence of collaboration, the defensibility of a search depends largely on an assessment of reasonableness. Specifically, did the searching party take reasonable steps to address the imperfections of the chosen search technology? Judge Grimm discusses the importance of utilizing best practices including quality assurance techniques, such as the sampling and evaluation of results, in order to demonstrate reasonableness, citing the Sedona Conference Best Practices Commentary on the use of Search & Information Retrieval Methods in E-Discovery as one source of best practices.

While the value of these approaches is clear, the reality is that both approaches are actually difficult to implement with traditional search processes and technologies. As a result, these approaches are infrequently followed today, leaving attorneys and litigation support professionals potentially exposed to the adverse outcomes discussed earlier.

cont. page 6

Limitations of "Black Box" Search

Why are these approaches so hard to follow in practice? Clearly, there are many reasons for this, least of which is the adversarial nature of litigation, which makes collaboration especially difficult. However, one of the primary reasons is the search technology itself. Today's search technology is largely based on software originally developed for enterprise search, whose purpose is to identify ESI across an enterprise for general business purposes. This technology has largely taken a "black box" approach to searches, where searchers submit queries one at a time and receive results with limited visibility as to how or why the query "found" the documents it produced. This approach makes sense in enterprise search where the user is chiefly interested in locating only the most relevant information as quickly as possible, but in the realm of e-discovery it creates a number of problems, including:

More over-inclusive and under-inclusive searches. All searches are prone to producing over-inclusive and under-inclusive results. "Black box" search technology makes it harder to reduce over and under-inclusiveness. A simple example of this would be a search for divers*, which is intended to find documents containing the words "diversity" and "diverse." This search will also find false positives, such as "diversion" or "diversification," creating over-inclusive results. With "black box" search technology, the person submitting the search won't necessarily know that this search will find these false positives and can't easily avoid these false positives. This limitation is often not a problem in enterprise search because the searcher often only cares about the top 10 documents. But, in e-discovery, all of these search results may need to be reviewed by attorneys at substantial cost.

Costly search sampling, analysis and refinement. Both Judge Grimm in *Victor Stanley* and Judge Baker in *In Re Seroquel* call attention to the value of refining searches through the use of best practice testing and sampling techniques. However, iteratively testing, sampling, and refining searches can be very time consuming and costly with today's "black box" search technology. Every search query needs to be refined one at a time and, in e-discovery, it is not uncommon to have to run hundreds of search queries.

Manual documentation of refinement process. One of the things that recent case law makes clear is the importance of documenting one's search methodology so that it can be presented if called into question. Judge Grimm, in particular, calls out the failure of plaintiffs "to provide the court with information regarding" their search methodology. With today's search technology, documenting search methodology is largely a manual task, making it costly and error-prone.

These limitations of today's "black-box" search technologies make following both the collaborative and best practices approach more difficult. Collaboration in e-discovery search is innately challenging because both parties are typically striving for different outcomes. One party often wants to find more relevant documents regardless of the cost incurred. Whereas, the other party wants to minimize the volume and cost of reviewing documents. Today, because the cost and time involved in testing, sampling, and refining searches during a "meet and confer" process is so high, the opposing parties typically negotiate search terms in a vacuum without the benefit of being able to evaluate the outcomes of each negotiating position. The resulting uncertainty and lack of visibility make collaboration much more difficult to achieve.

These same limitations also make it difficult to adopt a defensible best practices approach, where using refining and documentation best practices is essential for creating a "reasonable" search strategy. Because these best practices can be so costly and time consuming with today's "black box" technologies, they are often skipped, increasing the risk of adverse outcomes if they are ever called into question.

Transparent Search

In order to achieve more collaborative and defensible search practices, therefore, what is needed is for e-discovery search technologies to move away from "black box" techniques towards more "transparent" approaches that are optimized for the specific defensibility requirements of discovery. "Transparent" search technologies should be able to:

Minimize both under and over-inclusive results. Using the previous example, instead of finding documents containing "diversion" and "diversification," when a user enters a search for "divers*" a transparent search would provide the option to choose which matching terms to search for, including "diversity" and "diverse" and excluding "diversion" and "diversification." Transparency should also be used to optimize other "black box" search technologies including stemming, fuzzy, and concept searches.

Enable efficient analysis and refinement of searches. Transparent search technology should be capable of executing, analyzing, and refining large numbers of searches simultaneously instead of one at a time. It should be easy to sample both retrieved and non-retrieved documents.

Automate documentation of the refinement process. Transparent searches should provide automatic search reporting and enable the producing party to create an audit trail of the decisions that were made based on analysis and

cont. page 7

refinement of a search. This reporting capability should track which terms were selected for the search, and which results have been excluded for either collaborative or defensible best practices purposes.

Benefits of Transparent Search

The benefits of moving from a "black box" to a "transparent" search approach are significant. Transparent searches can arm attorneys and litigation support professionals with the ability to comply with emerging case law, and more cost-effectively follow either collaborative or defensible best practices approaches in their searching. Transparent searches can also reduce the cost of e-discovery by making culling during a search more effective, and reducing downstream production and review costs. While transparent search technology has powerful benefits, it is important to keep in mind that no technology is a "silver bullet" and that ultimately the value of a technology depends on how well it is used. Until science develops true artificial intelligence, the practice of search is likely to remain imperfect. In the interim, greater transparency can play an important role in reducing the cost and risks of this imperfect, but necessary technology within electronic discovery.

William Uppington *is director of product management at Clearwell Systems, a leader in intelligent e-discovery, and co-leader of the Analysis working group within EDRM. William blogs on topics related to e-discovery technology at http://www.clearwellsystems.com/e-discovery-blog/*