

Controlling Legal Costs – Service Providers

What's The Big Deal About Search?

Theodore Sedgwick Barassi

SYMANTEC CORPORATION

“For lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread.”

– Magistrate Judge John Facciola of the U.S. District Court of Washington, D.C., in *U.S. v. O’Keefe* (D.D.C. Feb. 18, 2008).

As suggested rather colorfully by Judge Facciola in the *O’Keefe* case, the use of search in support of electronic discovery creates potential pitfalls for attorneys. Lawyers need to think before they search; without careful attention to factual research, search results testing, and collaboration with opposing counsel, discovery search can become a fool’s game.

Judge Facciola followed-up on *O’Keefe* within a month in a second search-related opinion, *Equity Analytics, LLC, v. Lundin*, (D.D.C. 2008). His thinking was further elaborated upon several months later in a tightly reasoned 43-page opinion by Judge Paul Grimm in *Victor Stanley, Inc., v. Creative Pipe, Inc.* (D. Md., May 29, 2008). Collectively, these three decisions have forced litigators to confront the limitations of search as a discovery tool. They outline the dangers of ill-conceived and untested search strategies and call for greater cooperation between litigating parties to address discovery search methods early in the pretrial process.

The search testing and collaboration theme continued into 2009 with Judge Andrew Peck’s decision in *William Gross Construction Assoc. Inc. v. American Manufacturers Mutual Insur. Co.* (S.D.N.Y. March 19, 2009) in which he called for “careful thought, quality control, testing, and cooperation with opposing counsel in designing search terms or ‘keywords’ to be used to produce emails or other electronically stored information (‘ESI’).” These decisions have enhanced our understanding of the issues surrounding the application of search for the discovery of ESI, and memorialized broad practice requirements to which litigating parties can adhere.

Perhaps not unexpectedly, the decisions in these cases also unleashed some-

thing of a feeding frenzy amongst vendors selling advanced discovery search tools based on proprietary algorithms. It is not uncommon to hear vendors claim that keyword searches are intrinsically unreliable, or that it is impossible to construct reliable discovery searches without significant (and expensive) input from technical search experts. Terms for various search methods have entered the discovery lexicon – fuzzy search, algebraic search, probabilistic search, concept search, search results clustering, etc., and while many of these methods may have an important role to play in improving search efficiency and accuracy, they are often poorly explained or understood by the legal community that relies on them. Instead of helping to ease the problem, much vendor marketing has had the paradoxical effect of sowing fear, uncertainty, and doubt in an area in which the courts were hoping to clarify requirements and encourage practical, collaborative results. The issue is nicely framed in Jason Krause’s April 2009 article in the *ABA Journal*, “*In Search of the Perfect Search*,” quoting Bill Speros, a Cleveland-based e-discovery consultant:

As for the lawyers currently conducting e-discovery, they are finding themselves in an untenable position. “If we use some fancy search technology that we can’t explain, we’re put in harm’s way in front of the judge,” says Speros. “And if we use dumb and naive keyword searches, we’re in harm’s way. For lawyers like me, we just want to know one simple thing: What will work for me and my client?”

The stakes in how this question ultimately gets answered are not small. In large matters, the discovery of electronic information can often cost in the millions of dollars – so being able to reduce the amount of potentially relevant documents for review is extremely important. In addition, being able to find relevant data in the overwhelmingly vast sea of information on most organizations’ systems can be a real challenge.

So, what can litigants do to navigate the issues surrounding search? How do you construct search and retrieval strategies that have the track record of defensibility and ease of explanation offered by traditional keyword search while avoiding the pitfalls that befell parties in the cases cited above?

Although both technology and the law are likely to continue to evolve in this area, there are several identifiable themes that emerge in the judicial treatment of search issues. These themes are backed by research and analysis done by expert electronic discovery organizations such as the Sedona Conference, as well as the TREC Legal Track, operating under the U.S. Department of Commerce’s National Institute of Standards and Technology (NIST). The Electronic Discovery Reference Model (EDRM) Search Working Group has also recently undertaken a comprehensive treatment of discovery-related search, with the publication of its EDRM Search Guide.

At a high level these themes can be described as: putting in place good processes around the search method and

tailoring the technical search tools used within that method to work most effectively with the specifics of the matter.

The Role Of Process In Good Effective Discovery Search

The recent rulings indicate that the search method used must include both technical and non-technical processes sufficient to show that the search is likely to produce a reliable result. It is clear that key elements for reliability are not limited to the type of search technology selected, and it is certainly true that non-technical factors, such as thoughtful consideration of search terms, working with custodians who have a personal knowledge of the matter, testing for reliability, and working collaboratively with opposing counsel are crucial in constructing an effective methodology.

Research

As indicated most forcefully by Judge Peck in the *William Gross* case, many attorneys construct searches without taking the time to review the search details with key custodians having personal knowledge of the matter. Having input from custodians can help clarify whether preliminary search strings contemplated by attorneys can be improved by including elements requiring personal knowledge, such as code words or terms of art, for example. Key custodians, being subject matter experts on the underlying facts and circumstances of a particular case, can also provide the kind of context that will help defend the methodology employed for searching for the documentary evidence pertaining to that matter.

Testing

Internal testing of the relevancy of the chosen set of terms can also increase the defensibility of the method. Often, testing against a sample set of information known to contain responsive content, such as a custodian’s email archive, can help improve the efficacy of the method and enable the party to predict how the result will apply to the larger set of information to which the method will ultimately be applied. Combining well-researched and well-tested search strings will greatly facilitate the defensibility of the process, both with opposing counsel and the bench.

Collaboration

Collaboration with opposing counsel to validate the underlying assumptions of the scope, relevant terms, and search testing scenarios is an additional key element in getting to a defensible search and retrieval method. Clearly communicating these issues early on with specific and documented requests for feedback can also help show good faith and bolster credibility in front of the bench. Including the opposing side’s input as a secondary research and testing phase – again against a reference-able data set – can help encourage real collaboration, instead of meaningless point-counterpoint arguments relating to search details.

Technology And Understanding Technical Limitations

Parties need to be able to understand and communicate why certain search

methods are being employed as part of the overall search and retrieval methodology. As Judge Grimm wrote in *Victor Stanley* “the party selecting the methodology must be prepared to explain the rationale for the method chosen to the court, demonstrate that it is appropriate for the task, and show that it was properly implemented.”

Practically speaking, this means that the attorneys involved in the matter need to understand the scope and limitations of whatever method and tool they choose for the type of content for which they are searching. Certain search tools, for example, may rely on content indexing technologies that are designed to quickly identify and retrieve common data formats within larger volumes of data. Others may include indexing features to facilitate advanced querying techniques. While these features may be useful for internal research and testing against a subset of data, they may also result in huge, unwieldy indexes and over-production of content if used against the overall corpus of data included in the entire target collection set.

For testing results, too, the efficacy of the search method should be included. Testing the percentage of relevant search returns against a reference subset is important, but that test should also include an examination of what is being discarded within that reference subset. Because there is no specific accuracy benchmark that the courts apply to relevancy and non-relevancy thresholds for this type of sample testing (other than general reasonableness and reliability), it is important to also document with detail the level of testing done and prepare to defend that as a reasonable threshold.

Search And Information Management

Finally, it is worth mentioning that the best way to limit the risks and challenges associated with searching massive quantities of discoverable yet irrelevant content is to proactively reduce the information being created and stored. For email in particular, the implementation of a comprehensive information management program involving archive-based classification, retention, and, importantly, expiry can significantly reduce the volume of data that may be subject to costly and risky discovery searches.

In *Arthur Andersen LLP v. the United States*, 544 U.S. 696 (2005) Chief Justice William Rehnquist wrote: “Document retention policies, which are created in part to keep certain information from getting into the hands of others, including the Government, are common in business. ... It is, of course, not wrongful for a manager to instruct his employees to comply with a valid document retention policy under ordinary circumstances.”

It is clear that proactive information management has the support of the bench. If practitioners will take this to heart and begin to implement sound information management processes, they may lessen the need to invest in costly and complex search solutions. At the end of the day, these solutions are used to compensate reactively for significantly over-retaining content that should have long ago been thrown away.

Theodore (Ted) Sedgwick Barassi is Group Product Manager for e-Discovery and Information Risk at Symantec Corporation. He brings more than 15 years experience as an attorney to electronic discovery, records management and information security. He has practiced law as in-house counsel in financial services technology, electronic commerce, privacy and data protection. Both the Sedona Conference and the Electronic Discovery Reference Model have excellent resources publicly available on search. Further information can be found at EDRM <http://edrm.net/activities/projects/search> and Sedona Conference http://www.thosedonaconference.org/dlt-Form?did=Best_Practices_Retrieval_Methods_revised_cover_and_preface.pdf.

Please email the author at ted_barassi@symantec.com with questions about this article.