

InsideCounsel

Where angels fear to tread

By Clifford F. Shnier

September 5, 2008

In a previous column, "[A time to reap, a time to cull.](#)" I indicated that in a future column I'd go into greater detail about defensible search strategies. Unbeknownst to me at the time, U.S. Magistrate Judge Paul Grimm went into the kind of detail on this topic that will be hard to surpass.

Grimm's decision in *Victor Stanley Inc. v. Creative Pipe Inc.* 2008 (WL 2221841), along with *United States v. O'Keefe* 537 F. Supp., *In re Seroquel Prods. Liab. Litig.*, *Equity Analytics v. Lundin*, are all part of a recent line of cases that might be referred to as the "Angels Fear to Tread" opinions. In a nutshell, they put the brakes on the practice of simply brainstorming a keyword list to cull down electronic data.

As Judge Facciola stated in *O'Keefe*: "Whether search terms or 'keywords' will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics.... Given this complexity, 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**. This topic is clearly beyond the ken of a layman."

In *Victor Stanley*, much of Judge Grimm's opinion is contained in lengthy footnotes. One states: "Viewed in its proper context, all that O'Keefe and *Equity Analytics* required was that the parties be prepared to back up their positions with respect to a dispute involving the appropriateness of ESI search and information retrieval methodology—obviously an area of science or technology—with reliable information from someone with the qualifications to provide helpful opinions, not conclusory argument by counsel."

In other words, counsel can neither assert nor dispute the sufficiency of a keyword search on their own say-so. And pounding out a keyword list around a conference table in a half hour doesn't cut it, unless all parties have irrevocably agreed to that list.

E-Discovery consultants and service providers have seen keyword lists given to them by their clients that clearly were going to miss much of what apparently was being sought, and were going to hit much that was going to be irrelevant. As one consultant in this field commented: "The main thing to remember about the first keyword list counsel gives you is that it's always wrong."

Here's a generic example. Let's say one of the issues the case revolves around is dogs. The e-Discovery service provider is handed a list that includes the word "dog*". The attorneys making up the list knew they had to add the asterisk so they could get the plural form. (At least that much has been learned over the years.)

What they didn't realize is that it will also hit false positives on documents containing words like dogma, dogmatic and doggerel. So right off the bat, there will be some overinclusiveness. Also, if you're really interested in documents about dogs, you might need to come up with some synonymous words such as the word canine and maybe some words that are instances of particular kinds of dogs like the words beagle and golden retriever.

It's been pretty much well recognized for some time now in the literature and publications of thought-leading institutions such as the Sedona Conference that keywords alone suffer from the twin deficiencies of being both

over-inclusive and under-inclusive. In fact, this flaw was recognized in the 1985 study by Blair and Maron that stated keywords used in searching the optical character recognition full text of paper documents (there was no electronically stored information back then) were right only 20 percent of the time.

The initial keyword list should be treated as a good starting point. What is required is expansion and refinement of the initial keyword list by persons experienced and qualified to do it, and to iteratively test sample search sets of both hits and non-hits to get to a point where a particular expanded and refined search strategy appears statistically sound in hitting what is intended while omitting what is not to reasonable percentage of accuracy. It will never be perfect.

This sounds like a lot of expensive add-on time and labor to what was supposed to be the key to solving the information overload burden placed on the discovery stage of litigation. It doesn't have to be, and should not be viewed in that manner.

Judge Grimm continued: "The message to be taken from O'Keefe, Equity Analytics, and this opinion is that when parties decide to use a particular ESI search and retrieval methodology, they need to be aware of literature describing the strengths and weaknesses of various methodologies, such as The Sedona Conference Best Practices...and select the one that they believe is most appropriate for its intended task. Should their selection be challenged by their adversary, and the court be called upon to make a ruling, then they should expect to support their position with affidavits or other equivalent information from persons with the requisite qualifications and experience, based on sufficient facts or data and using reliable principles or methodology."

Software has existed for some time that automatically expand search terms into concepts, and/or generate synonym lists, and/or offer up words from the same stem that can be accepted or rejected, all with an aim to speed up this process. Provided the operator of this software knows it well, and documents all steps, and the software itself is able to show what it has done (that is, it is not a dreaded "black box"), such products can indeed be helpful. And the expertise of consultants skilled in this arena is your backup evidence to support what you've done.

Still, probably the most effective way to contain e-Discovery costs is to realize that one of the ways you can best serve your client is to work these matters out co-operatively with the other side.

As Judge Grimm observed: "[P]arties worried about the cost of employing properly designed search and information retrieval methods have an incentive to keep the costs of this phase of discovery as low as possible, including attempting to confer with their opposing party in an effort to identify a mutually agreeable search and retrieval method. This minimizes cost because if the method is approved, there will be no dispute resolving its sufficiency, and doing it right the first time is always cheaper than doing it over if ordered to do so by the court."

It may be counter-intuitive to our adversarial process, but co-operation with the other side may be the best thing you can do for your client. If nothing else, it may get you talking about settlement. A recent article in *The Economist* titled "The Big Data Dump—AA deluge of electronic information may overwhelm American civil justice" noted that now only two percent of cases go to trial. That's not necessarily a bad thing if they're settled for good reasons. As one of my clients back in my days of practicing litigation said to me, "If you have to go to trial, you've already lost, even if you win."

Reaching a settlement to save on discovery costs, especially electronic discovery, is not a good reason to settle. The source of the problem in discovery—technology having created massive volumes of information which would have been a fraction of that amount back in the days of paper—is also the source of the solution. Whatever else you may say about electronically stored information, one good thing about it is that it is searchable without having to do much at all to it. And therein lies the solution.

An integral actor in litigation support and e-discovery since the late 1980's, Clifford F. Shnier has been a

consultant, owner of his own service bureau and a senior sales and management executive with major e-discovery and litigation support providers. Prior to that, Shnier practiced law in Toronto for eleven years, litigating complex commercial matters, torts and criminal cases, with extensive trial and appellate courtroom experience. Based in Scottsdale, Ariz., since 1994, he is active in the e-discovery arena on both sides of the U.S.-Canadian border.

(c) 2008 *InsideCounsel*. A [Highline Media](#) publication. All rights reserved.