EDD Update

Electronic data discovery news and analysis

A joint project of Law Technology News and Law.com Legal Technology

LTN LAW.COM

July 23, 2009

Searching for the Definitive "Search" Standard

Raise your hand if you have a definitive answer on what is an absolutely defensible process for searching through a couple hundred thousand documents and produce all the relevant and responsive documents, and not include too many "false positives"? And don't even get me started with those pesky privilege documents (we'll leave that for another posting).

The Current Civil Discovery System

In the simple vernacular, our current system of civil discovery is whacked!

When talking about our search in our current system, I like to use the analogy of a field of haystacks. The number of haystacks we're dealing with depends on the size of the case. If we have a small case, we might be looking at only a handful of haystacks. We have a massive case, we can be looking at haystacks spread out over several acres. We start with the following facts: (1) most, but not necessarily all of these haystacks will have needles in them, though I can't tell you how many needles will be in each haystack or how many there are in total; (2) there are also two types of needles: Privileged and Non-Privileged. Now for the ground rules of the game: Your job is to go find every needle in every haystack, separate out and account for every "privileged" type needle by recording it in a log file, and then give that log file, along with all the non-privileged needles to the requesting party. If you fail to deliver every needle, you may be severely penalized. If your log file is incomplete or has too many errors, you may severely penalized. If you give the other side too many hay straws along with the needles, you may also be severely penalized. And by the way, there is a big ticking clock. Begin!

What's a poor farmer -- I mean lawyer -- to do?!

The Human Problem

Is the right answer to put together a team of contract attorneys to work their way through the documents one at a time conducting a first pass review? Why not? With law schools are pumping out more and more starving lawyers each year, the supply is so high that the street price to a vendor for providing a licensed contract attorney is approaching $40 per hour in mid-tier locations; and that's all in, including workstations in the vendor's space. At that price, it's like getting a super sale special of buy 1 and get 5 for free! The thing is, we're talking about (presumably) intelligent people who have made it through law school, passed a bar exam, and are being paid the societal equivalent of slave wages. Is this really a morale enhancing situation? As it is, studies have shown that humans tend to make MANY mistakes when performing subjective relevancy review. So how confident can we be in the ability of a bunch of underpaid, poorly incentivized humans to both find all the needles, separate out the privileged ones, and not leave us with a bunch of hay that we're going to get penalized for turning over?

The Machine Problem

On the other side of this conundrum is the suggestion that we throw everything at machines. If you ask a few vendors out there, they will be happy to tell you that the machines can run some amazing magic, and pop out the relevant documents with a high level of precision. But if we were to take big magnet and walk around each haystack, how confident can we be that got out every needle, which is in essence what we would need to do in defending our production?

The Statistical Sampling Model

In the middle, we have the growing movement of analytics and statistical sampling. Courtesy of a few friends on the Federal bench (Judge Facciola and Judge Grimm in particular), we are now working on the idea of using scientific methodology to test and validate the results of what a human did or what a computer did, and tell us how accurate they were to a X degree of confidence. The rub is that we have no ability to answer this simple question: what is a reasonable degree of confidence? 99.94% pure? Good luck! And to make things slightly more confusing, we must consider that the degree of confidence is a variable subject to proportionality and reasonableness, i.e., what's a reasonable degree of confidence for a case worth $500K, a case worth $5M, or a case worth $500M?

The Search Goes On

So the question remains, how do we come up with a definitive legal standard for conducting a reasonable search that is affordable, manageable and defensible in 94 district courts, never mind who knows how many state courts?

We have some great minds working on this question in TREC Legal Track, The Sedona Conference, and the EDRM Search project. But is all this work going to create a definitive process, or just more questions and more confusion? Moreover, what's the value of running these 3 major projects if there's not a simple, predictable, test that litigants can expect to be applied by a court?

Inquiring minds want to know!

By Eric P. Mandel at 01:15 PM in Commentary & Analysis , EDD Update Authors | Permalink | Comments (5) | TrackBack (0)

TrackBack

TrackBack URL for this entry:
http://www.typepad.com/services/trackback/6a00d8345280a669e201157228bbcf970b
Listed below are links to weblogs that reference Searching for the Definitive "Search" Standard:

Comments

Perhaps arrogantly, I raise my hand.

I think that you are looking for solutions in the wrong place.

The critical thing is the process. TREC is not designed to provide answers to your questions, jut tools. It would be swell if we could have a tool that we could pour ESI into and get out just the right set, but it simply won't happen. No search protocol is a substitute for thought, rather the technology and methods can amplify the value of the thought.

What does it take to have a defensible search process?

Honesty. Actually do what you said you were going to do. Many of the cases that appear to involve search issues, to my mind, actually revolve around questions of honesty.

Transparency. Be able to say what you are going to do and be able to say what you did do.

Effectiveness. Use technology and processes that actually work and be able to show that they worked, not just in the abstract, but in your specific case.

Reasonableness. Be proportional. Be cooperative. Spend your time, effort, and money where it will do the most good. Part of reasonableness is knowing where to set your standards. TREC, for example, found that human reviewers are not terribly accurate. I have written about this here: http://orcatec.blogspot.com/2009/04/trec-legal-track-2008.html. Two reviewers agreed with one another only about 72% of the time, meaning that if human review is taken as reasonable by definition, you don't have to be very good to exceed that standard. In the end, judges and attorneys are constantly making reasonableness judgments, the mere fact that these judgments have numbers associated with them does not change that.

Evaluation. Measure what you have done. You cannot know that you have done a good job if you don't measure. There are several points in the discovery process where you need to examine what you have done and determine whether it has been reasonable.

These are the main features of a defensible process. It is not mysterious and need not be vexing.

Posted by: Herbert Roitblat | July 23, 2009 at 05:18 PM
Your problem is a strawman: "That dog don't hunt," as we say in Texas. And the proposition of a "whacked" system may buy accolytes in a Sarah Palin-esque "we like the way it used to be but never really was" way, but it doesn't help.

The standard is not and never has been perfection. It's never been "absolutely defensible." The unfounded notion that if you don't get every needle or include a tad too much hay straw you will be sanctioned is specious and terroristic. It's the sort of huff-and-puff that's used to distract us from the fact that the efforts actually put forth are often so poorly conceived and -executed as to not stand in sight of competence, let alone perfection.

I don't know where the notion arose that because you can't achieve a perfect e-discovery effort, it doesn't matter how badly you do it. No one has been sanctioned for a diligent, good-faith EDD effort gone awry. The sanctions cases aren't a litany of "gotchas." Lawyers and clients in those decisions fell far short of reasonable competence, diligence or honesty.

When lawyers start consistently doing a reasonably good, cost-effective job on e-discovery albeit with some inevitable error, then we can waste time worrying about Platonic ideals. Until then, cursing the system without hard facts is not warranted or helpful. It's just a rant.

Posted by: Craig Ball | July 23, 2009 at 06:54 PM
Thanks for your comments. I would like to note that nowhere in my posting did I advocate the expectation of perfection. In fact, I agree that perfection is impossible. The issue I was trying to raise, and apparently without success, is that when it comes to search, the lack of uniform, justiciable standards as to what is "good enough" versus potentially sanctionable is creating an unreasonable hazard in the legal profession.

Right now, our community is doing wonderful work studying the sought after equilibrium of recall and precision; we are writing and publishing significant papers on search functionality and theory; and we are evaluating the application of algorithms and advanced mathematics to searches of large volumes of data subject to legal discovery. All of that is great!

But attorneys cannot simply be expected to "consistently [do] a reasonably good, cost-effective job on e-discovery albeit with some inevitable error" absent both education and predictable uniform standards that will be applied in courts across this country. Furthermore, judges -- many of whom do not at this time understand the technology -- cannot make reasoned determinations of fact and law without some test that incorporates the factors for what constitutes “good enough”.

In Zubulake, Judge Scheindlin provided us the framework for the analysis of what materials could be deemed as not reasonably accessible. Later codified in the FRCP, this standard is now filtering its way into the States. Do we not need the same consideration and thought put into developing a legal test as to what constitutes a reasonable search? For without such as standard, how can we ask a reasonably informed lay attorney to sign their name to a verification stating that the search methodology chosen and employed in their case was reasonably calculated to meet their client’s discovery obligations?

Many of us have discussed this issue in private. My hope that as a community we can discuss this, and other such issues, in this public forum while applying a level of civil decorum that is above throwing about judgmental buzzwords such as “Sarah-Palin-esque” “terroristic” and “platonic ideals”.

Posted by: Eric P. Mandel | July 24, 2009 at 05:07 PM
It seems to me that the answer lies somewhere between Erics call for standards and the responses by Craig and Herb. As Craig says , there's no such thing as a perfect process and as Herb says, it's ALL about the process.

In fact I made that very point back in February on my blog when I quoted John Martin: it's the archer not the arrow. (ttp://docnativeblog.wordpress.com/2009/02/18/its-the-archer-not-the-arrow/)

And really that's the problem to my mind. We get so lost in the technology we forget that we're not building a space shuttle here .. or a car or even a widget. We're searching for truth. Try to publish a standard for that.

OK and even if we're not looking for truth we're dealing with a process that is different every time. Different documents, different document types, different operating systems, different users ... you get the point.

To use Craigs analogy, you don't train a dog to hunt by just teachng him to do the same thing over and over. If you're duck hunting, he has to swim. If you're racoon hunting he needs to know how to tree. If you're fox hunting he has to work with other dogs. If you're bear hunting he has to know how to talk so he can tell you he's not THAT stupid even if you are.

But I do agree with Eric that we need to have more public debate over these issues. Otherwise the dogs are running the hunt and not the hunters.

Posted by: Tom O'Connor | July 26, 2009 at 02:22 PM
I'm surprised to see hostile reaction to this post. I agree with Eric that more specific guidance than the vague "reasonableness" standard we

currently have would be a welcome development.

On the other hand, I don't think we can (at least with current technology and methods) expect to find "a definitive search solution that is affordable, manageable and defensible in 94 district courts, never mind who knows how many state courts." There are too many variables to lay down a single protocol or perfect mix of methods that applies equally well in all cases. Doing so would be folly.

I do think we ought to provide more guidance to attorneys who don't, and never will, understand technology to the depth of the (e)discovery thought leaders. It's also folly for us to expect the entire profession will some day acquire that depth of knowledge in order to undertake (e)discovery with "reasonable competence [and] diligence." Those of us that do understand should seek to provide our less savvy brethren more specific guidance on what it means to conduct (e)discovery with "reasonable competence [and] diligence."

It seems that many people are already working on providing just that, including members of the judiciary. As with any area of law, I expect improved guidance via case law as we collectively figure things out. Eric's example of statistical sampling is a good one—recent cases have identified sampling (statistical and otherwise) as a good practice to improve the process, but that's vague. In the case of statistical sampling, should results be at the 90% confidence level or are we expected to apply the more rigorous 95% level (usually reserved only for medical research)? Similarly, when we sample, what error rate is acceptable? Are different error rates acceptable for samples of different haystacks? I expect it won't be long before we get some case law on these points.

Posted by: Chris Spizzirri | July 27, 2009 at 11:17 AM

Verify your Comment

Previewing your Comment

Posted by: |

This is only a preview. Your comment has not yet been posted.

Post Edit

Your comment could not be posted. Error type:

Your comment has been posted. Post another comment

The letters and numbers you entered did not match the image. Please try again.

As a final step before posting your comment, enter the letters and numbers you see in the image below. This prevents automated programs from posting comments.

Having trouble reading this image? View an alternate.

Continue

About Incisive Media | About Law.com | Customer Support | Privacy Policy | Terms & Conditions

Copyright 2009. Incisive Media US Properties, LLC. All rights reserved.

An incisivemedia website