

The Metropolitan Corporate Counsel

www.metrocorpcounsel.com

Volume 16, No. 9

© 2008 The Metropolitan Corporate Counsel, Inc.

September 2008

E-Discovery, Analytical Tools – And Human Judgment

The Editor interviews Gene Eames, Senior Consultant, Data Analytics, SPi Legal.

Editor: Please tell us about your background.

Eames: I worked as a litigation support manager in a New York law firm for 17 years, dealing mostly with document control databases, case management and general practice support systems. Over time I developed quite a bit of expertise in technology, particularly text management and searching tools. In 1997 I moved on to become a technology consultant where I specialized in building unstructured data retrieval systems and knowledge management systems for a number of global corporate entities. After seven years working in that environment I decided that my experience in litigation support enhanced by my experience working in corporate IT enabled me to bring value to the e-discovery space.

I joined SPi four years ago, originally as a project manager on a very substantial e-discovery case for a large pharmaceutical company, and I subsequently moved into the data analytics practice that was being created at SPi.

Ken Shear and David Kittrell mentored me on their approach of using data analytics to achieve targeted and validated results when attempting to reduce large collected data sets into more manageable review collections. David and Ken are founding fathers in the e-discovery marketplace. Both have worked extensively at e-discovery vendors since the industry first began to take shape. Their shared experiences and thoughtful consideration gave birth to the data ana-

lytic approach which SPi considers one of its core offerings.

Editor: What are automated data analytic tools?

Eames: They can actually be a number of different things, although people tend to group them all together. I think what most refer to as data analytics are tools that will in an automated fashion organize, categorize or tell you something about your data. Many of them work by using statistical or mathematical analysis of words that are in documents to derive some meaning that may not be readily apparent.

Editor: Are these analytic tools simply search engines?

Eames: Not necessarily, although they may include search functionality. Most people in the legal world understand what full text searching is because that is what Westlaw, Lexis and Dialog have done for years. Full text search tools take words in a document and index them, creating a record of word locations which permits you to quickly find a word on its own or in proximity to other words. This allows one to quickly find search “hits” without the need to read through each line of text. Data analytics tools are something different. Instead of simply recording exactly where in a document each word lives, these tools consider the words in relation to each other to derive some contextual meaning. This can allow a tool to take a set of docu-



Gene Eames

ments and put them in smaller groups by subject matter in a highly automated fashion. The underlying technology is usually very complex and difficult to explain, so most users don't know exactly how it works. They just see that it sort of works. That in my view can potentially be a problem, depending upon how and where one applies the tools. If you have to explain to a judge how you selected responsive data, you better be able to explain how you did it. However, if you use the tool to simply classify your review sets for speedier review, then I don't think it matters how you did it, as long as everything that should have been reviewed is reviewed.

Nonetheless, such tools can play an important role in selecting a review set. Because they represent possible defensibility issues we take care to use them in such a way that we won't have to defend how they work. We translate what we learn about data into something that's much more transparent. For instance, in the testing of a search term we take samples of data and use analytic tools to give us an idea of what's in the data. We then take what we learn to craft more effective text searches knowing people have a better grasp of that technology.

Editor: I would assume a lawyer would give you the search terms for the purposes of searching?

Eames: That is correct. And if one is not applying an approach similar to SPi's, then that may be the end of the story. Lawyers often dream up search terms in a vacuum because they are not looking at the documents. They know about the case and the issues and based on that they

Please email the interviewee at g.eames@spi-bpo.com with questions about this interview.

develop the terms for which they wish to search. Those searches are submitted to a search engine and you get what you get, but the bottom line is that text searching can be a really inaccurate and imprecise thing. There are many studies out there that will tell you that. However, those studies do not say that you cannot apply text searching effectively – they just say that it’s kind of difficult to do, and that is where testing and validation should come into play.

Editor: So, can any automated tool accurately select responsive discovery documents?

Eames: It might be able to, but I don’t know that I would rely on it without validating it, anymore than I would rely upon running a blind text search without looking to see whether it got the expected results. Additionally, regardless of how effectively the tool retrieved what you intended, it may not correlate to subjective review calls made by an attorney. Even with validated search terms a search hit does not necessarily translate into a responsive document. To meet the requirement of reasonableness one should strive to target a review set with a high percentage of actual responsive documents while limiting the review of irrelevant documents. That goal can be met by validating your results through non-hit sampling to make sure you are comprehensive enough, and by doing review call feedback to make sure you are not over-inclusive. Human judgment assisted by technology!

Editor: Could one of the reasons for poorly targeted review sets be that there might be alternate terms that the lawyer had not thought of?

Eames: Exactly, that’s very common, and we take steps to avoid that while still keeping the results from becoming overbroad. Many of the automated analytics search tools will automatically expand your terms to related terms or concepts. That may or may not be a good thing when you’re talking about litigation support. You may not want all the related terms because it may result in false hits to review and prove to be too costly. A basic tenet of our approach is to validate what you are doing by testing it and sampling results, so to the extent that we use

automatic tools to expand searches, we do it selectively in a human-assisted way, incorporating human judgment into the process to ensure effective application of any tool.

Let’s assume that there are many documents in a collection that do not have hits based upon current searches – even though the searches have been tested and revised through a series of iterations. We take representative samples of the documents containing no search hits and ask attorneys to review them to see if there are any responsive documents. If the attorneys find responsive documents, we know that we have to go back to the drawing board and craft revised searches. We repeat the process until counsel feels that we’ve taken all the reasonable steps that we need to take.

Editor: What do you and your colleagues at SPi bring to the table?

Eames: I have 25 years of experience using advanced technologies for unstructured text searching and structured database systems. Most lawyers do not have that level of experience. Seventeen years working in a law firm has given me an understanding of the lawyer’s world, and my time spent in corporate IT has given me a perspective on some of the difficulties of e-discovery from the corporate clients’ standpoint. I am not a lawyer – so while I am not hired to make legal decisions about the completeness of any discovery effort, I can assist attorneys in making good decisions on the technology front and give them some insight into what might be considered reasonable. My colleagues have similar backgrounds.

Editor: Will introduction of human judgment into the equation make judges more likely to find that a company’s e-discovery efforts are sufficient?

Eames: There is a lot of nuance in language and there are many ways that responsive documents can slip through the cracks. However, when I’ve spoken to judges about this, they remind me that they are requiring reasonableness – not perfection. Assume a court were to ask my client “What steps did you take to ensure that these search terms were the right search terms?” I feel we could take

the court through a process clearly demonstrating our due diligence, and the courts are looking for that diligence. The recent decision *Victor Stanley Inc. v. Creative Pipe Inc.*, 2008 WL 2221841 (D. Md. May 29, 2008) involved search-based privilege identification, and was not specifically about the selection of documents for review. However, I suggest the same arguments regarding search apply. In that case, Judge Grimm stated that simply using a search engine without testing its results was not good enough.

In the alternative, we show the evolution of the original search terms using statistical hit analyses we prepare of search results. We point out changes made in the search terms based upon what we discover from samples of the resulting documents. We show that our process typically brings us through multiple iterations of searches before reaching a comfort level. When we find something responsive during non-hit review, we indicate changes in the search terms necessary to supplement our review with such documents, and we continue the process until we stop finding responsive documents in the non-hit sets – much of which may ultimately not be reviewed.

“We incorporate what we learn into an overall process of data selection which includes testing, validation and human judgment with the goal of being reasonable in our discovery responses.”

Thus we avoid being forced to say we have some tool that somehow makes everything right, but we can’t explain exactly how the magic button works. Instead we are able to say that we use a number of tools, some of which involve complex technology, and that these tools help us learn about our documents. We incorporate what we learn into an overall process of data selection which includes testing, validation and human judgment with the goal of being reasonable in our discovery responses. We present the steps we follow, along with documentation of the results of those efforts.