BAR BULLETIN

Playing the Smart Odds:

A Probability Model for Proportional, Efficient E-Discovery

By Larry G. Johnson

You can see the power of the laws of probability at work in Las Vegas: those garish towers of gambling owe their existence to the odds that inevitably favor the house. In like manner, all the lawyers in a case can enjoy house odds if the principles of probability are applied intelligently to the discovery of electronically stored information ("ESI").

No, we're not talking about rolling the dice in the hope relevant evidence will magically appear. What makes an e-discovery probabilistic model discussed here work is your expert knowledge about the issues in the case, the context(s) within which those issues arose, and the kinds of evidence you can assume to be of importance if found.[1]

In other words, what makes the sea of available ESI not a crapshoot but rather a manageable mass that can be gleaned through culling and deduplication is the use of a feature in some of the text-searching software used by e-discovery experts. Such software not only isolates likely evidence through carefully constructed search terms, but it also can rank and weigh those search terms according to the relative importance you assign to each of them.

Let's get specific with a hypothetical case where an e-discovery probabilistic model is employed. Let's say your case involves claims of employment discrimination. The plaintiff alleges she was passed over for a promotion given to a less-qualified male, and she also alleges there has been an overall upper-management acquiescence in sexist slurs and inappropriate jokes directed at her that constitute a hostile work environment.

In a case like that you can bet the plaintiff's attorney is going to want to conduct discovery that will include co-worker text messages and voice mail from smartphones, as well as the emails from company and personal email accounts of those persons the plaintiff has reason to believe contributed to the alleged discrimination.

The plaintiff's attorney will likely send immediately a "preservation letter" reminding the defendant company that it has a duty to preserve all potentially relevant evidence, including ESI, and to establish a defensible litigation hold. That letter can become the trigger for attorney "meet and confers" under Rule 26(f) — negotiations regarding just what sorts of evidence must be preserved; when and where Rule 26(a) initial disclosures and document production should take place; and the crafting of a Rule 26(f) discovery plan (which, as far as I can tell, only a few savvy lawyers take full advantage of).[2]

It is at this important initial phase of discovery that concepts of probability theory can be brought to bear to the benefit of all parties in the matter, and such concepts are also consistent with newly emphasized proportionality requirements in recently revised Fed. R. Civ. P. 26(c) and state equivalents.[3] The days of the kitchen sink and turning over every rock in discovery are clearly over.

You build the foundation for the probabilistic model when you use your brain and instincts to decide who the key witnesses are for your case, and what you think the issues are. Discovery does not begin with a blank slate. The most critical component toward a probabilistic approach to discovery of ESI is common sense.

In our hypothetical, not all of the plaintiff's co-workers are going to be likely sources of relevant evidence. The plaintiff will most likely know from direct experience who some of the bad guys are and where key documents are likely to be found. So, the first stage set in the discovery plan is to have key witnesses' smartphones[4] produced, along with their emails to and from each other during a relevant timeframe.

After limiting first efforts to the key witnesses,[5] your Rule 26(f) discovery plan can articulate a carefully crafted, limited set of search terms that each side wants employed as it perceives the issues. Unanimity on search terms is neither realistic nor needed. Again, for economy and proportionality, each side should agree for this initial discovery phase to, say, 15 search terms.

In addition, each party should rank in order the search terms it wants employed so that the terms can accordingly become prioritized for potential relevance; that way, the responsive documents will be

sorted from most likely relevant at the top to least likely relevant at the bottom. In other words, there exists software that can take your search term rankings and employ artificial intelligence to sort the set of documents responsive to searches according to a percentile likelihood that they are relevant. A document scoring 99 percent is probably highly relevant; anything with a relevance ranking below 15 percent is almost certain to be irrelevant — I have found "irrelevance" can kick in with documents having a ranking even as high as 40 percent.

My use of such software has dramatically reduced the size of review sets for the final phase of attorney review for privilege and relevance. I use rankings of 15 percent or less as a conservative relevance cutoff point. I can justifiably cull those from the attorney document review set and keep them stored in separate folders in case my methodology is challenged by any party. For quality assurance, I do look at a sampling of documents in the 15-percent-and-under relevance population just to test my assumptions, and invariably these documents are false positives.

In a recent case, this probabilistic procedure reduced the number of potentially relevant documents responding to search terms from 96,720 documents with "hits" to 3,081. The cost savings in reduced attorney review time were enormous, and harried associates doing the document reviews were grateful for the reduced burden.

Is it possible that you may miss key documents when using a probabilistic model? Of course that is possible. But missing key documents is, in my experience, most often due to the inherent limitations of document searches, whether done manually[6] or with e-discovery software tools. And then there is the frequently elusive nature of language itself. Evil often occurs in mists. The "smoking gun" email that you could easily miss might say nothing more than, "OK, let's do it."

Exhaustive attempts to find that sort of smoking gun email short of luck would probably exceed the proportionality requirements of Rule 26. Lawyers have to accept that every now and then the house loses on a turn of the wheel — but playing the smart odds in e-discovery will make you a winner more often than not. Probably. ∎

Larry G. Johnson is a lawyer in Newcastle, and has been a member of the Washington bar since 1974. He recently served on the E-Discovery Subcommittee of the WSBA Escalating Cost of Civil Litigation (ECCL) Task Force. Besides being a litigator, for the past 20 years he has served as a consultant and expert witness in e-discovery matters. He does

business as Electronic Data Evidence (www.e-dataevidence.com).

[1] See my article in the January 2016 issue of the KCBA Bar Bulletin, "Bits, Bytes and Brains: Still the 'No-Brainer' E-Discovery Tool," where I argue your legal expertise and intelligent hunches about the issues and evidence far outweigh any currently available e-discovery technologies: *https://www.kcba. org/newsevents/barbulletin/BView.aspx?Month=01& Year=2016&AID=article6.htm.*

[2] See my article in the June 2016 issue of the KCBA Bar Bulletin, "Your Electronic Discovery Plan: Rule 26 – Hidden Nuggets for the Savvy Litigator:" *https://www.kcba.org/newsevents/bar bulletin/BView.aspx?Month=06&Year=2016&AID= article20.htm.*

[3] See "New Federal Rules Aim to Promote Proportionality in Discovery." at *https://www.hunton. com/files/News/95a86cf8-e593-4c14-8210-8b5e ab82435c/Presentation/NewsAttachment/a3ce6cea- 9f97-41a9-b965-8c49b85bc548/new-federal-rules- aim-to-promote-proportionality-in-discovery.pdf.*

[4] See my article in the October 2015 issue of the KCBA Bar Bulletin, "What's in Your Pocket? Smartphones: A New 'Safe Harbor' for E-Discovery:" *https:// www.kcba.org/newsevents/barbulletin/BView.aspx ?Month=10&Year=2015&AID=article1.htm.*

[5] Remember, discovery does not have to be done in one fell swoop; you can do it in iterative stages, where the results of one set of discovery requests can inform and narrow what you ask in subsequent ones.

[6] Even painstakingly slow and inefficient manual reviews of documents are no guarantee of success: "… a recent Price Waterhouse study … conducted a search of over 10,000 documents looking for a specific topic, author, and data range. A manual search of paper files performed by a staff of paralegals took 67 hours and found 15 documents. The same search, performed using a document management system, required 4.5 seconds and found 20 documents. The time differential between 67 hours and 4.5 seconds is compelling, but many corporate managers will be more concerned about the five documents that were missed in the manual search." *http://tdan.com/ capture-the-value-of-document-management-find- the-mess/5107*