



SEARCH, FORWARD

Will manual document review and keyword searches be replaced by computer-assisted coding?

First, there was manual review — the “traditional” method of document review. As a young associate at a major New York law firm in the late 1970s, I reviewed boxes of files for relevance, “hot documents,” and privilege. To gather the paper documents, you went to the client and asked where they kept files about “X” (“X” being the issue(s) involved in the lawsuit). Often there was a central file labeled “X,” and employees kept their own working files as well. Occasionally, you had to go to the dreaded warehouse, where boxes might

not be indexed, and working conditions always were less than ideal.

Review was linear. There was no way to deduplicate documents or organize them by types. You reviewed whatever box landed on your desk; colleagues might be reviewing a carbon copy of the same file. Hopefully, you both coded it the same. (Even today, it is not unusual for a document to be produced while another copy is on the privilege log.)

When associate billing rates became too high, firms turned to paralegals, staff attorneys, or contract attorneys.

Whether this had any effect on the quality of the review was beside the point; economics drove the change.

Despite its flaws, many senior lawyers (and some clients) still consider manual review to be the “gold standard” against which other review techniques are compared. While the volume of electronically stored information (and concomitant expense) has largely eliminated manual review as the sole method of document review, manual review remains used along with, for example, keyword screening. Let us consider whether manual review as the gold standard is myth or reality.

Two recent research studies clearly demonstrate that computerized searches are at least as accurate, if not more so, than manual review. Herb Roitblatt, Ann Kershaw, and Patrick Oot, of the Electronic Discovery Institute, con-

cluded that “[o]n every measure, the performance of the two computer systems was at least as accurate (measured against the original review) as that of human re-review.” (“Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review,” *Journal of Am. Society for Information Science & Technology*, 61(1):70-80 (2010).)

Likewise, Wachtell, Lipton, Rosen & Katz litigation counsel Maura Grossman and University of Waterloo professor Gordon Cormack, using data from the Text Retrieval Conference Legal Track, concluded that “[T]he idea that exhaustive manual review is the most effective — and therefore the most defensible — approach to document review is strongly refuted. Technology-assisted review can (and does) yield more accurate results than exhaustive manual review, with much lower effort. (“Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review,” *Richmond J. of Law & Tech.*, Vol. XVII, Issue 3, 1-48 (2011).)

Grossman and Cormack note that “not all technology-assisted reviews . . . are created equal” and that future studies will be needed to “address which technology-assisted review process(es) will improve most on manual review.”

KEY WORDS

Because the volume of ESI has made full manual review virtually impossible, lawyers have turned to keywords to cull ESI (particularly e-mail) for further (manual) review. A basic problem is that absent cooperation, the way most lawyers engage in keyword searches is, as Ralph Losey suggests, the equivalent of “Go Fish.” The requesting party guesses which keywords might produce evidence to support its case without having much, if any, knowledge of the responding party’s “cards” (i.e., the terminology used by the responding party’s custodians). Indeed, the responding party’s counsel often does not know what is in its own client’s “cards.”

The problems with keyword search

are well known. Lawyers are used to doing keyword searches in “clean” databases, such as Westlaw and Lexis, which use full sentences, full words (not abbreviations), and largely the same words to describe the same concept. E-mail collections are not clean databases. People use different words to describe the same concept; even business e-mails are informal, rampant with misspellings, abbreviations, and acronyms.

The object of search is to produce high recall and high precision. Recall is the fraction of relevant documents identified during a review, i.e., a measure of completeness. Precision is the fraction of identified documents that are relevant, i.e., it is a measure of accuracy or correctness.

When keywords return false positives — documents that have the keywords but are not relevant — the responding party has to use expensive manual review to find the truly relevant documents. It is not uncommon for a poorly chosen keyword to return more “junk” than responsive documents, i.e., low precision. The goal of search is to produce high recall and high precision (in a cost-effective way).

How effective is keyword searching? In 1985, scholars David Blair and M.E. Maron collected 40,000 documents from a Bay Area Rapid Transit accident, and instructed experienced attorney and paralegal searchers to use keywords and other review techniques to retrieve at least 75% of the documents relevant to 51 document requests. Searchers believed they met the goals, but their average recall was just 20%. This result has been replicated in the TREC Legal Track studies over the past few years.

Judicial decisions have critiqued keyword searches. Important early decisions in this area came from magistrate judges John Facciola (District of Columbia) and Paul Grimm (Maryland). See *United States v. O’Keefe*, 37 F. Supp. 2d 14, 24 (D.D.C. 2008) (Facciola, M.J.); *Equity Analytics, LLC v. Lundin*, 248 F.R.D. 331, 333 (D.D.C. 2008) (Facciola, M.J.); and *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 260, 262

(D. Md. 2008) (Grimm, M.J.).

I followed their lead with *William A. Gross Construction Associates, Inc. v. American Manufacturers Mutual Insurance Co.*, 256 F.R.D. 134, 136 (S.D.N.Y. 2009) (Peck, M.J.).

“This Opinion should serve as a wake-up call to the Bar in this District about the need for careful thought, quality control, testing, and cooperation with opposing counsel in designing search terms or ‘keywords’ to be used to produce e-mails or other electronically stored information (‘ESI’),” I wrote.

My opinion concluded: “Electronic discovery requires cooperation between opposing counsel and transparency in all aspects of preservation and production of ESI. Moreover, where counsel are using keyword searches for retrieval of ESI, they at a minimum must carefully craft the appropriate keywords, with input from the ESI’s custodians as to the words and abbreviations they use, and the proposed methodology must be quality control tested to assure accuracy in retrieval and elimination of ‘false positives.’ It is time that the Bar — even those lawyers who did not come of age in the computer era — understand this.”

Despite these (and other) judicial criticisms of the use of keywords without sufficient testing and quality control, many counsel still use the “Go Fish” model of keyword search. Cooperation is important, but without testing and quality control cooperation alone is not the answer.

COMPUTER-ASSISTED SEARCH

Even with keyword searching, lawyers have turned to certain computer-assisted approaches to further reduce review cost. Boolean connectors can be used (such as “and,” “or,” “w/in,” “but not”). In addition, deduplicating the ESI (either within a single custodian or across the entire production) greatly reduces both volume and the chance of the same e-mail being coded differently by different reviewers. Grouping “near duplicates” takes that a step further. Threading e-mail chains is another

useful technique.

If the hot topic in 2010 conferences was proportionality, this year it is computer-assisted coding, often generically called "predictive coding." By computer-assisted coding, I mean tools (different vendors use different names) that use sophisticated algorithms to enable the computer to determine relevance, based on interaction with (i.e., training by) a human reviewer.

Unlike manual review, where the review is done by the most junior staff, computer-assisted coding involves a senior partner (or team) who review and code a "seed set" of documents. The computer identifies properties of those documents that it uses to code other documents. As the senior reviewer continues to code more sample documents, the computer predicts the reviewer's coding. (Or, the computer codes some documents and asks the senior reviewer for feedback.)

When the system's predictions and the reviewer's coding sufficiently coincide, the system has learned enough to make confident predictions for the remaining documents. Typically, the senior lawyer (or team) needs to review only a few thousand documents to train the computer.

Some systems produce a simple yes/no as to relevance, while others give a relevance score (say, on a 0 to 100 basis) that counsel can use to prioritize review. For example, a score above 50 may produce 97% of the relevant documents, but constitutes only 20% of the entire document set.

Counsel may decide, after sampling and quality control tests, that documents with a score of below 15 are so highly likely to be irrelevant that no further human review is necessary. Counsel can also decide the cost-benefit of manual review of the documents with scores of 15-50.

To my knowledge, no reported case (federal or state) has ruled on the use of computer-assisted coding. While anecdotally it appears that some lawyers are using predictive coding technology, it also appears that many lawyers

(and their clients) are waiting for a judicial decision approving of computer-assisted review.

Perhaps they are looking for an opinion concluding that: "It is the opinion of this court that the use of predictive coding is a proper and acceptable means of conducting searches under the Federal Rules of Civil Procedure, and furthermore that the software provided for this purpose by [insert name of your favorite vendor] is the software of choice in this court." If so, it will be a long wait.

Judicial decisions, including *Victor Stanley, O'Keefe* and *Gross*, are highly critical of the keywords used by the parties. These decisions did not "endorse" or "approve" of keyword searching. Nevertheless, lawyers seem to believe that the judiciary has signed off on keywords, but has not on computer-assisted coding.

In addition to reluctance to be the guinea pig for a decision on predictive coding, lawyers perhaps are concerned that they will have to go through a *Daubert* hearing as to the "admissibility" of the results of predictive coding. Perhaps this fear comes from *O'Keefe*, where Judge Facciola said that opining on what keyword is better "is truly to go where angels fear to tread," and is a topic "beyond the ken of a layman and requires that any such conclusion be based on evidence that, for example, meets the criteria of Rule 702 of the Federal Rules of Evidence," dealing with expert opinions.

Lawyers' fears in this regard seem largely misplaced. First, Facciola's comments were directed at keywords, but everyone is using keywords, and I know of no decision after *O'Keefe* requiring expert testimony as to the use of keywords.

Second, with due respect to Facciola, I do not think *Daubert* applies — it applies when an expert will testify at trial in order to admit into evidence opinions or results (e.g., the result of DNA testing reveals a match).

Here, the hundreds of thousands of e-mails produced are not being offered into evidence at trial as the result of a

scientific process. Rather, whether the handful of e-mails offered as trial exhibits is admissible is dependent on the document itself (e.g., whether it is a party admission or a business record), not how it was found during discovery.

That said, if the use of predictive coding is challenged in a case before me, I will want to know what was done and why that produced defensible results. I may be less interested in the science behind the "black box" of the vendor's software than in whether it produced responsive documents with reasonably high recall and high precision.

That may mean allowing the requesting party to see the documents that were used to train the computer-assisted coding system. (Counsel would not be required to explain why they coded documents as responsive or non-responsive, just what the coding was.) Proof of a valid "process," including quality control testing, also will be important.

Additionally, counsel can point to the TREC study and other reported studies that generally show that computer-assisted coding technology works at least as well if not better than keywords or manual review.

Of course, the best approach to the use of computer-assisted coding is to follow the Sedona Cooperation Proclamation model. Advise opposing counsel that you plan to use computer-assisted coding and seek agreement; if you cannot, consider whether to abandon predictive coding for that case or go to the court for advance approval.

Until there is a judicial opinion approving (or even critiquing) the use of predictive coding, counsel will just have to rely on this article as a sign of judicial approval. In my opinion, computer-assisted coding should be used in those cases where it will help "secure the just, speedy, and inexpensive" (Fed. R. Civ. P. 1) determination of cases in our e-discovery world.

Andrew Peck is a United States magistrate judge for the Southern District of New York. E-mail: Andrew.Peck@nysd.uscourts.gov.