

PREDICTIVE CODING | A SPECIAL REPORT

It's been more than two years since a federal magistrate publicly endorsed the use of predictive coding in electronic discovery, and even though many lawyers still haven't heard of the technology, most of the heat has gone from the debate. Whether you call it predictive coding, technology-assisted review or computer-assisted review, the term connotes use of computer algorithms to search for dispositive evidence. In this special report, we asked litigation experts to take stock and examine some of the techniques out there.



ISTOCKPHOTO/BLACKACK30

Tools Let Attorneys Follow the Breadcrumbs

Analytics applications can help litigators identify surface patterns that point to the key evidence.

BY GARETH EVANS AND DAVID GRANT

The extraordinary recent technological innovations—in particular, the explosion in information volumes along with the computing power and analytics applications to make sense of that information—stand ready also to revolutionize how we search for and analyze evidence in litigation.

Powerful tools now exist that not only allow counsel to review and classify large volumes of information in an expeditious and relatively cost-effective manner (e.g., predictive coding), but also to do the important early detective work of ferreting out documents that can make or break a case (e.g., various other analytics applications). To be sure, only a small minority of litigation attorneys are using such tools. The question, however, is not whether legal professionals will adopt them, but when. In the meantime, those who do so



ISTOCKPHOTO/GENFERBRUARY

may gain substantial advantages for their clients.

Predictive coding has already begun to quietly gain traction. We say “quietly” because, despite all the vendor hype, it appears that few who are using predictive coding are broadcasting that fact. Although lawyers are rarely thought of as revo-

lutionaries, Gil Scott Heron's poem about the last era of great societal change, the 1960s, is not entirely inapt: “The revolution will not be televised, Brothers; The revolution will be live.” Change has arrived. You just may not be hearing a lot about it yet.

Predictive coding is essential-

PREDICTIVE CODING

ly supervised machine learning designed to reproduce knowledge based on what it has learned from the training documents with which it has been provided. It works by replicating information it knows—i.e., the relationship between the features of the training documents to the coding calls provided for those documents—across the remainder of the document population. It can be particularly effective for the review of large volumes of documents.

Predictive coding can be used to prioritize likely responsive documents, to replace manual review of nonresponsive documents or to replace the use of keywords. Of course, the traditional key words and human review still works well in most cases—there is a reason why it remains the standard approach to document review—and it is the responding party's right to determine appropriate methodologies and technologies to identify and produce relevant and responsive documents.

Although predictive coding is a powerful tool that can (and should) be considered for many cases, other analytics tools with additional important strengths may be useful as well. We have found that a combination of one or more of these analytics tools and predictive coding can be effective in first finding important documents and key facts and in gaining important insights about the overall document population and individuals involved (analytics), and then efficiently undertaking a comprehensive review of the document set (predictive coding).

In contrast to predictive coding's application of a set of rules across a population, analytics applications—such as concept clustering and visual analytics—are designed to organize documents based on their internal content and their similarity to, or dissimilarity from, other documents in the set. By letting the documents speak for themselves such analytics tools can surface patterns—a trail of “bread crumbs” that you can follow to the relevant facts. Particularly when the relationship and similarity of individual documents to each other can be ascertained, thus unearthing relatively small “pockets” of interesting documents.

Visual analytics provides this information in the form of graphically rendered clusters or pockets of documents that you can click through to more detailed levels and even the underlying documents themselves. It enables smart human researchers to leverage the visualized information and ignore most of the documents while zooming in on the key documents.

In many ways, it is the perfect complement to predictive coding. Whereas predictive coding uses rules from a small set of documents to categorize the entire set, visual analytics uses the content of the entire set to let a researcher zoom in on the key fact documents quickly. Documents found through the use of analytics can be used as a “seed set” to help train a predictive coding tool.

This fact-finding paradigm is new, enabled by advances in large-data visualization. The technology is so powerful that the easiest mistake is

to assume it's the same as comprehensive review. The goal, however, is precisely not to thoroughly look at everything that might be producible (a process driven largely by your opponent's formal document request strategy), but instead to follow the breadcrumbs to the key-fact documents. A team of researchers operating under a fact-finding paradigm, with the freedom to use their good judgment (and operating separately from the main comprehensive review team) can make very effective use of these tools.

The beauty of using an analytics tool up-front is that it changes the game—instead of the production process gradually unearthing the story, the elements of the story can be unearthed up-front and used to drive the strategy. Used together, analytics tools and predictive coding can provide a powerful “one-two punch” in identifying important documents quickly and then getting through the review of a large document population in an efficient and cost-effective manner.

Gareth Evans is a partner at Gibson, Dunn & Crutcher and co-chairman of its electronic discovery and information-governance practice. David Grant is a senior managing director at FTI Consulting Inc. and leads its discovery consulting group.