

# The Metropolitan Corporate Counsel®

www.metrocorp-counsel.com

Volume 20, No. 12

© 2012 The Metropolitan Corporate Counsel, Inc.

December 2012

## E-Discovery Steps Outside Of The Black Box

*The Editor interviews Joe Looby, Senior Managing Director, FTI Technology.*

**Editor:** Tell us about your professional background.

**Looby:** I've had a very interesting career. After graduating from law school, I joined the Navy and became a Navy Judge Advocate (JAG) stationed across the country handling courts marshal prosecution defense, investigations and environmental work. After four years in the Navy, I joined the New York State Environmental Conservation Department, where I handled matters regarding cleanups of the Hudson River, the St. Lawrence Waterway and New York Harbor.

Having become interested in technology along the way, I built and programmed an enforcement tracking system at the Environmental Conservation Department that was used by New York State's enforcement professionals to track and follow through on the environmental violations they were detecting. I then got into consulting, which I've been doing for 12 years – the early years building computer models, including a patented system for detecting fraud in financial statement audits and then nine years at FTI – handling matters such as leading the technology investigation of Bernard Madoff. I've been working with our research and development team at FTI on our software and technology, including the predictive coding solution that we're going to talk about today.

**Editor:** FTI recently conducted a survey on predictive coding adoption. Were there any surprises?

**Looby:** There were. We spoke with approximately two dozen counsel at Fortune 1000 companies and attorneys at AmLaw 200 firms, and they gave us some very useful qualitative and quantitative data on a wide range of topics related to predictive coding. Half the respondents we spoke to had used

predictive coding, most of them through pilot or experimental programs. And of those, over 90 percent had a positive experience with predictive coding and expect to use it more often in the future.



**Joe Looby**

In terms of specific use cases, respondents described using predictive coding to cull out relevant documents and prioritize documents for review. They also noted that the major adoption inhibitor to using the technology was the “black box” phenomenon: that the process is not transparent. This underscored that there is some confusion in the market. Some providers claim their product is not a black box because you can see what comes out of it. But users are concerned with what's happening *inside*. So I developed an article (which can be found at <http://www.fticonsulting.com/global2/critical-thinking/fti-journal/predictive-coding.aspx>) actually explaining the underlying math and process inside the machine to clarify how the technology works.

**Editor:** What are some of the key takeaways for corporations considering the use of predictive coding?

**Looby:** Predictive coding is not a silver bullet; it is just one tool in the toolbox, and attorneys do not appear to be ceding all control to the machine. No one (yet) sends a document out the door without review for privilege or relevance just because the machine has predicted it will be responsive.

Respondents overwhelmingly agreed that humans become an even more important part of the process. At the beginning of the process, you select a training set of documents that you give the computer so it can “study” and “learn.” If you make mistakes in creating that training set, the computer will then extrapolate from there, leading to

the amplification of an error in the larger population as those initial mistakes are repeated over and over again.

Finally, there are certain matters predictive coding does not lend itself to. For one, while there's no bright line, most respondents felt that predictive coding was unnecessary in matters involving 100,000 documents or fewer. Second, predictive coding is not effective for “needle in the haystack” searches because computers can't predict what they weren't trained on; they're incapable of imagination and they can't think outside of their box.

**Editor:** What is FTI's offering?

**Looby:** Ours is a full-service solution that combines our people, our process and our technology because we think that all three are required for the process to be defensible.

On the people front, our team of data science experts works with counsel to review the training set, coding the documents first for responsiveness and second as potentially privileged. Then, in our technology, the computer model studies and learns from that, then applies what it's learned to the larger collection and assigns responsive and privileged scores to the document collection. The higher the score, the more likely it's responsive. We've got a lot of other visualization technologies that can then be used to data mine and review what comes out of the computer model.

The last part of the process – and we think this is a key differentiator – is that FTI has many different business segments, one of which is the economics segment. Working with a top-three statistician in the world, we built a statistical process into the workflow, and we check the key performance indicators of the model – recall and precision – in a multi-step process using statistical sampling and verification by counsel and experts. We think it's critical to have that information and to document it, so that a few years down the road if you are challenged on your review, we'll stand behind it.

*Please visit [www.ftitechnology.com](http://www.ftitechnology.com) for more information about FTI.*

**Editor: Does part of the training that you mentioned refer to training the people who train the machine?**

**Looby:** There's an element of that. We think this is best done with a small team. Say there's a project at the company – "Project X" – and there's some patent dispute around Project X. Probably there's an associate general counsel inside the corporation serving that business unit who has a wealth of knowledge about Project X and who can also protect privilege. A senior attorney at a law firm is likely involved because federal rules require an attorney to sign the discovery response representing that it complies with the law, and it is generally a proper and reasonable response. Additionally, a subject matter expert can work with one of our data scientists to train the model to review the training set. This small group will develop the instruction and protocol for what's responsive and what's non-responsive based upon their understanding and their review of the incoming discovery requests from the other side or from the regulatory body.

**Editor: How is this different from other predictive coding offerings?**

**Looby:** First, we offer a complete solution. You don't have to buy the software and then find a data scientist skilled in running the software. Second, ours is not a black box solution. As we explain in the article, the technology uses elementary school mathematics – like addition and subtraction. The machine takes the training set and extracts all the words and phrases, and then it assigns each word and phrase in the training set a numerical score.

I think this is an elegant technology because it learns the way we all do – by trial and error. It picks one document at a time and "guesses" if it's responsive or nonresponsive. If the machine guesses incorrectly, it records and tries to learn from that mistake by building up a table of words and phrases that it has learned are important and a table of words and phrases that are unimportant, creating a frame of reference from which the machine can make predictions.

Other technologies actually take the whole training set into memory to create a model. Ours doesn't; it handles one document at a time, and that enables it to more easily scale to larger matters.

Last but far from least is our expert team. We've got world-renowned statisticians and PhD economists with credentials and decades of experience who can clearly explain the process to users.

**Editor: How do you ensure defensibility?**

**Looby:** There are two methods. First is an expert statistical quality assurance to ensure the coding accuracy. When rating information retrieval, we use "recall," which refers to false negatives – i.e., the model incorrectly classifies some responsive documents as nonresponsive – and "precision," which refers to false positives – i.e., incorrectly classifying some non-responsive documents as responsive – as the key performance indicators. We perform a series of tests throughout the process to measure performance to the desired level of confidence.

The second step is a kind of quality control through a visual verification of what comes out of the back of the computer model. We take that large pool of responsive documents and put it into our Document Mapper visualization tool. This clusters together similar documents, and using colors to denote coding, we can hone in on any clusters with inconsistent codes. This second level of verification, and the use of visualization technology, is unique within the industry.

**Editor: What are some of the common use cases lawyers should consider for predictive coding?**

**Looby:** The first and most important one is to cull out the irrelevant documents, which is a far more defensible process than using the traditional Boolean search method. A Boolean search may be great for online shopping, but it's hardly the optimal tool for discovery.

The second one is prioritizing documents and getting those documents with higher relevance scores to the key review team more quickly.

The third use is testing human review, in which documents are reviewed by people and scored by machine. The results from the human review teams can be tested against those from the machine, and where there are wide variants, the team can investigate those discrepancies to improve the overall quality of the review. More responsive documents may also be discovered in the process.

**Editor: What are the costs associated with predictive coding?**

**Looby:** In our survey, most respondents couldn't give a cost or savings amount for predictive coding, in part because they were using pilot programs. There may have also been a lack of transparency about how

the predictive coding they'd used was priced. Some providers use per-click or -gigabyte fees while others do not – the market is still figuring out how to price this new service. Here at FTI, it's a low-cost service based on the consulting our data scientists do with the outside counsel and/or subject matter experts in helping that team train the model.

**Editor: What are the potential cost savings with predictive coding?**

**Looby:** They are tremendous. We've done some economic modeling that reveals that identifying a responsive set using our predictive process and then doing some manual review based on that set can be done for a fraction of the cost of conducting a fully manual review.

**Editor: But predictive coding is just part of the story, right? How do lawyers effectively review the remaining documents?**

**Looby:** This is key. As I mentioned earlier, no one is (yet) using predictive coding and then producing relevant materials without at least one round of attorney review. So how can legal teams review the remaining materials quickly and cost-effectively? To answer this, FTI provides a full toolbox of analytics software to help legal teams review and hone in on important materials. For example, I mentioned Ringtail's Document Mapper feature earlier. It can help reviewers reconcile any predictive coding discrepancies and enable reviewers to quickly review highly relevant materials. Other tools enable legal teams to quickly focus in on other key data points, such as particular custodians or time frames.

In total, we provide a complete toolbox to meet a client's information retrieval and legal review goals.

**Editor: Is there anything else you'd like to add?**

**Looby:** I'd like to highlight a recent case with several interesting challenges. First of all, the company had a lot of data – nearly two terabytes – and that had to be collected, reviewed and produced to the government in a condensed time frame. The data was in the U.S. as well as Europe – which means data privacy issues came up – but through our predictive coding technology and the deployment of our mobile investigation teams on the ground, we saved the client about a million dollars. This speaks to the power of predictive coding, as well as the breadth of offerings FTI provides to clients.