

The Rise of Analytics in E-discovery:
**A Lifesaver When You're
Drowning in Data**



There were just over **295,000 civil suits filed in U.S. district courts in 2014, an increase of 4 percent over 2013**. On top of that, there were another 81,000 criminal cases (including fraud) filed.

The good news is that the number of civil filings rose just a little from September 2013 to September 2014, and the quantity of criminal filings dropped to the lowest total since 1999.

The bad news is that examining and producing evidence to pursue or defend these suits (or to develop a settlement proposal) have become an almost inhumanly complicated and time-consuming business — nor is it inexpensive.

In short, legal teams today are faced with a growing evidentiary nightmare as the number of emails, text messages, spreadsheets, PowerPoint presentations, PDF files and other digitized business documents rises dramatically every day. And as the lines between business networks and social networks narrow, the variety and complexity of data also are expanding and now include audio, video and other forms of information such as sensor data from Internet-connected devices.

But despite this avalanche of information, courts have been slow to change rules or timelines for cases. Even as the volume of structured and unstructured data continues to grow, court cases must proceed, and attorneys who fail to meet their court-ordered deadlines risk hefty fines in addition to assuming the increasing costs associated with the e-discovery process.

This is no small problem.



The Pressure Is On

Humans (even those with a law degree) no longer can sit behind a computer screen or, worse, around a conference room table groaning with documents to determine what is pertinent or privileged, relevant or immaterial. There's no question that technological tools — what is called e-discovery — are required to sift through it all.

It's not unusual for attorneys and their teams to have to examine tens of thousands of emails and other documents in an effort to turn up perhaps a dozen or so key pieces of evidence around which to build a defense or settlement strategy.

Getting to the necessary evidence, finding the key documents and eliminating the irrelevant ones are costly, as well as time-consuming, tasks. The majority of Fortune 1000 corporations now spend an estimated \$5 million to \$10 million annually on e-discovery, with several companies reporting expenses as high as \$30 million in 2014. **A full 70 percent of these costs were tied directly to the physical review of documents**, according to a 2012 RAND study. **That boils down to about \$1.8 million per case or approximately \$18,000 a gigabyte.**

These expenditures will only go higher. That's because **by 2020, the data we create and copy annually will reach**

44 zettabytes (44 trillion gigabytes), according to research organization IDC. Global mobile data alone are projected to reach 52 million terabytes in 2015, an increase of 59 percent from 2014.

And it is on this sea of data that legal teams must set sail to discover the key themes and critical documents for each matter — from investigations to a wide variety of litigation events.

Today, technology, and particularly the use of data analytics, is what separates the winners from the losers in the costly game of beat the clock to meet court-imposed discovery deadlines.



The Problem with Keywords

Most of the time, attorneys don't know at the outset exactly which documents will be useful to a case. Traditional review strategies during the discovery phase of litigation often entail identifying search terms likely to locate responsive documents in the data set. Called keywords, these terms are developed after researching the issues at hand and interviewing individuals associated with the matter. And while these keywords can be useful and help frame the review process, they have serious limitations when used alone.

Suppose attorneys suspect that a dishonest executive is expunging

incriminating documents from a company database. To try to prove it, the attorneys will conduct a keyword search of email and documents using terms such as “delete,” “erase” or “kill.”

Ultimately, the search is limited to those terms obvious to the attorneys. But criminals can be devious, which makes using these traditional methods to find relevant patterns in data difficult if not impossible.

In this way, text-based e-discovery searches limit what can be uncovered. For example, in the famous Enron data set — rich and diverse with lots of so-called “noise” (making it ideal to use for testing e-discovery technologies) — Enron executives used many code words (often “Star Wars” references) to disguise illegal activities. These code words would have provided attorneys with a whole armory of smoking guns that could have been used to reveal a host of crimes and misdemeanors; that is, if the attorneys knew what those words were. What reasonable attorney would have thought to use “Millennium Falcon” or “Chewbacca” in a keyword search of an energy company’s transactions?

New e-discovery tools, however, can recognize patterns and alert attorneys to the occurrence of seemingly inexplicable (and, therefore, attention worthy) words and phrases.

For that reason, the largest companies now are racing to adopt ever more sophisticated e-discovery tools, as are legal firms. Recent market research conducted by LexisNexis among 125 Am Law 200 law firms indicates that one-third of these firms were looking to invest in e-discovery technologies in 2015. According to the research, **38 percent said they would change their document review platform, 30 percent planned to change their production tool and 28 percent of respondents said they were changing their processing software.**

But small to midsized companies, including most law firms, already are overwhelmed — struggling to keep up. And costs are mounting all the time.

And as **70 percent of e-discovery overhead is associated with the review phase** (as opposed to processing or producing documents), the better the technology and ability to use analytics, the more quickly a firm can drive down costs. Improving that technology — and doing so without incurring crippling expenses — is critical to the survival of small and midsized firms.



How Small and Midsized Firms Can Compete

One promising approach to shrinking both the time and charges associated with e-discovery — and thereby allowing law firms and companies with fewer resources to stay in the game — is the emergence of data analytics and visualization software that can help lawyers organize and analyze mountains of data in new and different ways. Designed to reveal trends and focus a legal team’s review efforts to provide more efficiency, visualizations and dashboards can accelerate the discovery of key facts and, consequently, the development of case strategy while reducing time and costs. And since these analytics are built from keywords found in the electronically stored information itself, they can eliminate an attorney’s expensive guesswork (fishing expeditions) by (in effect) allowing the documents to describe themselves to the legal team.

Returning to the case of the executive suspected of deleting incriminating documents, attorneys came up empty handed using keyword search terms like “delete” and “erase” and other obvious analogs. However, by using analytics technology, the lawyers were able to identify and review alternative words, phrases and concepts and make connections across them. This is because analytics identifies all the concepts in a body of documents, not just those words, phrases or concepts that lawyers think may be important.

In the case of the crooked executive, the analytics tool surfaced the repeated use of the word “obliterate,” which, in fact, turned out to be the word the executive used to identify information he wanted deleted.

Essentially, what analytics technology does is find themes, recurring words, phrases or concepts automatically, without any keyword input. It presents a highly navigable framework that enables lawyers to see what they may not have considered (or had any reason to consider) at the beginning of the e-discovery process.

For example, along with identifying the repeated use of the word “obliterate,” analytics technology tied the repetitions to a particular time frame, which so often is a challenge for legal teams. Nailing down a specific time period can help pinpoint other parties that may have had knowledge of the suspected wrongdoing — for instance, parties that received emails with the word “obliterate” or used it themselves in their own communications. This provides the answer to one of the most critical questions involved in any case: Who knew what, and when did they know it? Answering that can open up an entirely new or different set of data collections for examination (such as the emails of other executives overly fond of the word “obliterate”), plus it gives researchers multiple entry points to huge data stockpiles.

An additional benefit of combining data is that this allows lawyers to identify and then explore smaller and/or unique subsets of data that are most relevant to the case and to do so quickly. One example might be emails exchanged among certain executives during the time frame when a large number of documents were added to or deleted from a database.

All these capabilities work to accelerate the development of a case strategy and can have a dramatic impact on negotiations with opposing counsel, saving valuable time in the race to meet court-imposed discovery deadlines and, of course, lowering costs.

Another high-level feature of e-discovery analytics systems is a technique called predictive coding, which especially can be useful in cases that have extremely tight deadlines or include massive data sets. Predictive coding works by employing sophisticated algorithms to select a subset of all case documents — say, 1,000 out of a universe of 10,000. The attorney reviews those 1,000 documents to determine their relevancy. The analytics system then analyzes that set of 1,000 — essentially learning from it — to identify and code key trends or patterns in the other 9,000 documents automatically.

This document sampling and machine learning process can reduce enormously

the time required to analyze a given data collection. Using traditional keyword search methods on the same universe of documents under a tight time frame would take a team of attorneys weeks to complete. And, correspondingly, during those weeks, the charges mount. Analytics technology can shave a significant amount of time and expense from this process, with accuracy rates between 95 percent and 100 percent.

Predictive coding is invaluable in establishing the scope of which documents need to be included in the discovery phase. In a patent infringement case involving two companies, a prosecuting attorney, for instance, might ask for all emails exchanged by every engineer at the company suspected of infringement. By using predictive coding, attorneys for the defendant might be able to identify as relevant a set of emails between just two engineers who worked on the patent over a period of nine months and, as a result, argue to limit the required documents to these emails

only. This work makes producing the documents faster and helps a company meet court-ordered discovery deadlines. Analytics also affords attorneys the opportunity to formulate their strategy earlier in the legal process, which allows them to determine much more quickly whether the evidence warrants a fight to the bitter end or a proposal to settle.

Even better, e-discovery analytics software today often has visualization capabilities that allow it to present data graphically. A visual approach to analytics can help attorneys rapidly determine what is and what isn't relevant to a case. Legal teams can visualize trends, summarize data, see multiple decision points, and drill down and out of data quickly and dynamically to identify an issue's key factors. In fact, visualizing data in lawyer-friendly ways is one of the most compelling advances in analytics software used in e-discovery — and a great boon to firms with limited or constrained financial and human resources.



Staying in the Game

To remain competitive, all large law firms and companies need to expand their e-discovery tools to include those with analytics capabilities, but it especially is important for smaller and midsized law firms. Doing so will require a concerted effort to overcome a profession-wide legacy of technology aversion and the legal profession's age-old preference for paper. Currently, less than 10 percent of law firms of all sizes are using analytics in an efficient way, according to FTI Consulting estimates.

But as e-discovery tools become more and more lawyer friendly and as visualization proves its worth, this situation is changing. Happily, analytic capabilities increasingly are being integrated or bundled into e-discovery software platforms, as opposed to the former practice of selling them separately under different licensing and pricing arrangements. This, of course, helps keep costs in check.

The bottom line is that the benefits of using analytics far outweigh traditional e-discovery methods, reducing both the time and expense of preparing a case. Today, for small to midsized law firms particularly, it may very well be the key to their survival. ■

JR Jenkins

Senior Director
Technology
FTI Consulting
jr.jenkins@fticonsulting.com

For more information and an online version of this article, visit ftijournal.com.