Regaining Control of ESI: Get Discovery Back on Track

By Rich Turner, VP Marketing, Content Analyst Company

The exploding volume of electronically stored information (ESI), and the increasing pressure to preserve it and be able to produce it quickly, is a major challenge for litigation teams. While the wide availability of ESI as evidence makes it a key aspect in virtually any litigation matter, the reality for many is that the expense, human resources and required adherence to federal regulations associated with producing this are overwhelming and projects are chronically undermanned.

There are a number of companies offering solutions or developing products that promise to speed e-discovery. Despite the progress being made by these companies, the sheer growth of ESI continues to outpace the improvements these solutions offer. Many solutions also require a major change in tools and processes to be effective—time and resources simply not available in the highly competitive litigation market.

Rather than wholesale, disruptive change, there is another option available today, that is more of a "course correction" and is easily incorporated into existing e-discovery workflows. Providers of next-generation technologies and review platforms are bringing the much talked-about power of "conceptual search" engines to the e-discovery field, because conceptual search eliminates issues involving keywords, and yields responsive documents regardless of specific wording. The corollary of conceptual search is "conceptual categorization," which is a greatly overlooked tool that can be easily incorporated into existing e-discovery processes today without any major transitions or specialized training.

Conceptual categorization uses small sets of examples or key documents and can quickly sort through massive collections of documents and identify only those documents that are similar or related to the provided examples. The power of conceptual categorization is manifold: it is a very fast process, able to handle hundreds of GBs of information in rapid order; easily scaled; easily repeated; and very defensible. By implementing conceptual categorization at the start of the ESI processing workflow, providers have found they can quickly identify and either set aside or eliminate volumes of irrelevant, non-responsive information that would otherwise consume valuable time and resources only to be discarded later in the process.

Let's look at the current process being used today by most solution providers. Consulting firms and ESI software solutions typically base their multi-step processes around a combination of manual and automated filtering, removing duplications and formatting of the information. This traditional approach consumes a great deal of time gathering immense amounts of data which are then loaded into a system that searches, processes or codes the data. The result is an e-discovery process that fills a central database with volumes of overly broad and irrelevant data that often is converted to non-native formats, losing metadata and the ability to retroactively search that data.

Let's contrast this to the more streamlined approach outlined in this article. Rather than an entire collection of data being run through multiple processes, it automatically goes through a categorization engine as the initial step—without being manually de-duplicated or culled. By identifying all potentially relevant documents out of the collection first, the possibility of missing something is greatly reduced. It then enters the usual workflow, following whatever tools and programs already being used. The difference is that a much smaller set of data—and much more relevant—is the product being processed.

This is made possible because a conceptual search algorithm is used for this categorization—essentially a superset of appropriate keywords along with contextual meaning. In the past, relevant documents were often not identified because of missed or inadequate keywords. This distinction is very important as courts are pressing for inclusive strategies rather than exclusive ones.

Conceptual categorization is a very fast process that doesn't add any significant time to the overall workflow. By only subjecting potentially relevant documents to further processing, the workflow is far more efficient than culling the whole raw collection as an initial step.

An additional benefit to this approach is its repeatability. To further add to the problems with ESI, it is often provided or given in "chunks" and in no particular order. The process outlined above is easily repeated as additional documents are released. The client can begin his or her review of relevant information, perform redactions, etc., in parallel with the production efforts rather than having to wait for a comparison with previous ESI. ▮

Content Analyst's Solution

Content Analyst has been in the search optimization industry in the government and private sector for 11 years. Our latest solution, CAT, has been implemented by a number of forward-thinking legal businesses. Our partners in the legal field have found that the current serial process of collecting, culling, extracting and producing ESI for use in litigation is a bottleneck in the overall management of e-discovery. They have implemented Content Analyst's conceptual categorization at the front of the process, thereby making a small—but significant—change to the traditional workflow for early case assessments and ESI production that yields tremendous benefits.

Rich Turner is vice president of marketing for Content Analyst Company; he is a veteran of the software industry and recognized for implementing vertical market strategies at emerging and advanced technology providers.

Content Analyst Company is a provider of advanced search and document analytics software; headquartered in Reston, VA, their public sector efforts have been providing key technology to government and intelligence agencies for more than a decade, while their commercial sector work has successfully focused on solution providers in the legal, publishing and content management markets. They can be reached at 1-888-349-9442, or info@contentanalyst.com.