



Pre-Discovery Analytics

Early Case Assessment and Pre-meet & Confer

By Nicholas Croce



As the amended Federal Rules of Civil Procedure (FRCP) begin to take hold in the litigation process, courts, plaintiffs and defendants are charting new courses in the management of electronic discovery. In the first nine months since enactment of the FRCP, there have been several rulings that indicate a new willingness by the courts to define “reasonableness” in the identification and collection of potentially responsive discovery information. Concurrently, technology for rapid collection and pre-review analysis has made significant advances, enabling enhanced early assessment of electronically stored information (ESI) and strategic analysis to pre-determine risk and prepare for the newly mandated meet & confer conferences. This white paper discusses the application of concept searching and analytic tools to provide insight into electronic data early in a case and possibly reframe the early litigation questions, “How do we gain strategic advantage?” and, ultimately, “*Should we litigate?*”

New Tools Provide Strategic Information Earlier in Litigation Process

Until now, lawyers were forced to answer the question of “Should we litigate?” with limited access to information. Before the advent of electronic information, paper-based documents were often hard to access and sort through for pre-discovery and early assessment. Unless critical documents were easily obtainable from key custodians, lawyers were forced to defend the early parts of the case with little knowledge as to the worth and cost of the effort.

With the proliferation of ESI over the past 15 years, information has become even more difficult to access. Documents, both paper and electronic, have exploded in quantity as well as complexity. Litigants on both sides were often unsure as to what kinds of documents existed and whether to include them in the discovery process. Execution of preservation and litigation holds became blanket orders to hold everything from file cabinets and hard drives to backup tapes and live email databases. The collection and processing of this information was (and still is) a Herculean task, and there was little to no opportunity to conduct early assessment when faced with the difficult and time consuming task of electronic discovery.

The recently amended FRCP recognize the challenges of finding and managing ESI and are focused on the reasonable management of the process so that cases can be decided on merits, rather than threat and cost. The new requirements to include ESI and the introduction of meet & confer conferences to discuss electronic data preservation and production are already affecting the discovery process. At the same time, new second-generation software technology has come forward to assist lawyers with the overall management of ESI for risk and cost control. Information can now be monitored in real time, preserved and collected on a semi-automatic basis, processed more easily with faster normalization of data and reviewed on an accelerated basis with new conceptual searching tools. In the face of the rising tide of information and increased pressure by the courts to respond in a timely manner, we are witnessing a sea change in the collection and review process, potentially saving corporations millions of dollars in litigation costs.

Key examples of new technology include active email archiving, rapid collection and conceptual searching. Active monitoring and compliance solutions for email are promising to detect and prevent the “smoking guns” from ever being sent. These solutions have potential but will take several years to fully implement across large enterprises. Moreover, as quickly as these solutions are incorporated, new communication technologies, such as integrated voice messaging and remote devices, will increase the challenges in deploying active solutions. Rapid collection technologies also hold significant potential to target and pre-process (cull and de-duplicate) much of the data at the earliest stages of litigation. Finally, concept searching and analytics can now be applied early in the process, yielding strategic information that can help defendants and plaintiffs alike to understand evidence and decide whether to pursue the case, and, if so, develop a strategy for managing the litigation process.

How Analytics Can Assist in Early Stage Litigation

Although analytics are relatively new to litigation and electronic discovery, conceptual analysis tools have demonstrated a strong ROI by accelerating the linear document review process. Analytics clustering of like documents in groups has been proven to speed review by a factor of five or more, with the immediate effect of lowering attorney review costs. While the value of analytics in the traditional review process is established, constraints on scale and search of large datasets have limited the use of certain tools for strategic uses.

As new products for collection and analytics have come to the market, another powerful application of concept searching is coming into focus: the ability to find key documents across large datasets at a much earlier stage in the litigation process. If we consider applying analytic technologies earlier in the thought and planning process, they can play a key role in helping to find the critical documents in early assessment as well as be applied to the testing of concepts and keywords for use in meet & confer negotiations. The analytical power of tools like Inference can help litigators gain a significant advantage in the litigation process and insight into data very early in the course of an action.

Early Case Assessment

At the early formulation of a law suit, attorneys discuss the merits of the case with their clients and focus on the risks associated with entering into litigation. This process is based on limited exposure to details and facts of the case. With the continually developing maturity of technologies used for electronic discovery and corporate data management solutions, key data is now often readily available and with reasonable access costs. Even before litigating, counsel knows *who* the key custodians in a potential matter include. Counsel also has insight into the date range of *when* issues concerning a matter would be in question. Should attorneys proceed with the filing, there would ultimately be volumes of additional data requested by the defendant with a tremendous cost impact to produce.

By taking a sampling¹ of data of the key custodians through the period of time when the action is suspected, critical insight into the facts of the case may be revealed. When concept searching analytics are applied to this data and the population is clustered by concept, information can be prioritized, allowing counsel to focus on critical documents of the case within only a few days. This process can provide much deeper insight into the facts, highlighting critical factors that dramatically impact both case strategy and the ultimate filing of a case. If damaging documents are found, counsel can decide to settle and save the time and costs of extended discovery and litigation. If the case has merit to proceed, the party that has utilized analytics will be able to plan litigation strategy and be better prepared for the ensuing discovery and production phases.

¹ Although a strategic decision at this stage, the use of sampling is reasonable, permitted under the FRCP, Rules 26(b)(2), 26(b)(2)(C), and 34(a)(1).

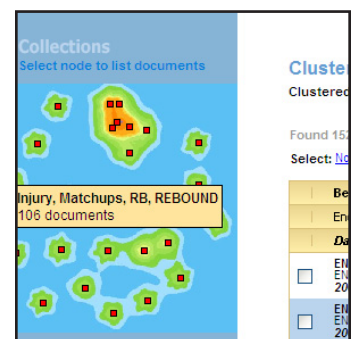
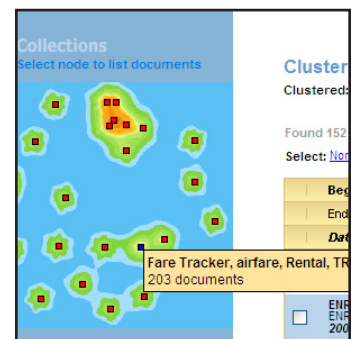
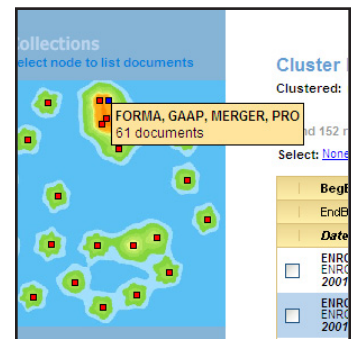
Pre-meet & Confer Preparation

Once a case has been filed, the amended FRCP mandate strict timelines for meet & confer conferences. These conferences address multiple negotiations for ESI, including the agreement of “keywords” to be used for limiting the production of ESI. For the foreseeable future, keywords will continue to be considered an acceptable method for producing ESI, although progress is being made in accepting analytics and concept searching as a valid method for producing information. As noted by U.S. Magistrate Judge John M. Facciola, “Once restored, how will they be searched to reduce the electronically stored information to information that is potentially relevant? In this context, I bring to the parties’ attention recent scholarship² that argues that concept searching, as opposed to keyword searching, is more efficient and more likely to produce comprehensive results.”³

Although keywords and date ranges are currently the basis for ESI production and review, the use of concept searching and analytics can be leveraged to provide fast and timely insight into how keyword negotiations can benefit a case. Without prior analysis of electronic data, critical meet & confer negotiations often happen on a blind basis. By design, plaintiffs hope to cast a wide enough net (within the legal boundaries) to find key pieces of evidence, while also pressing the defendants to incur significant time and expense to trigger settlement discussions. It is not unusual for defendants to request fewer than 100 keywords, while plaintiffs request thousands.

How can counsel effectively argue for search terms and keywords without any information? Is a keyword damaging to the defense and should they strive hard to have it removed? Or is it a giveaway that will send the plaintiffs off on a wild chase? To answer these questions and negotiate keywords, it is imperative for counsel to have better knowledge of what is in the data.

Through the use of analytics, counsel is able to see issues of importance or irrelevance on a rapid and targeted search basis. Shown at right by example using the Inference tool, analytics have the power to take a set of documents and group them by concept. Without any prior knowledge, input by users or previous review of the documents, the analytic tool highlights and groups documents, enabling quick identification of key issues and further insight into the corpus of discovery data.



As noted above, Figures 1–3 highlight potentially responsive and nonresponsive document sets in Enron’s case, with issues including acquisitions, travel information and sports. Inference’s cluster mapping provides a visual representation of conceptual groupings and interrelations of documents found across the dataset.

² George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?* 13 Rich. J.L. & Tech. 10 (2007).

³ U.S. Magistrate Judge John M. Facciola, *Disability Rights Council of Greater Washington v. Washington Metropolitan Area Transit Authority*, WL 15855452 (D.D.C. June 1, 2007).

In these examples of the Enron litigation database⁴, the analytics system surfaces potentially hot documents under GAAP and acquisitions (Figure 1), while also identifying non-responsive document sets related to travel reservations (Figure 2) and sports (Figure 3). Hot documents can be forwarded for pre-meet & confer analysis and non-responsive data, including those identified through “Show Similar,”⁵ that can be segregated for later, low-level review. Through this workflow process, counsel will have a much clearer picture of the keywords that are responsive as well as the strategic insight to include or exclude keywords. Defendants will potentially be able to limit the scope of discovery, saving review costs, as well as negotiate more forcefully on sensitive keywords.

Going Forward—Better Information Earlier

The convergence of the new FRCP for managing electronic discovery and new software technologies are providing counsel with the opportunity to gather strategic information at much earlier stages of litigation matters. Extending beyond their established use for accelerating linear review, concept searching analytics can provide significant insight for both early risk assessment and meet & confer negotiations. Counsel can now better define the merits of a case as well as manage risk and cost through the discovery process.

About the Author

Nicholas Croce, President of Inference Data, has led the creation and development of Inference, the company’s next-generation analytics software for electronic discovery. Prior to joining Inference, Nick was the president of Doar Litigation Consulting, a leader in electronic discovery and courtroom technologies.

About Inference Data

Inference is a leading e-discovery software for analytics-driven assessment and cost-effective review. Highly scalable and user intuitive, Inference incorporates the Autonomy IDOL™ engine to prioritize datasets, find key documents and accelerate the review process. For more information, visit www.inferencedata.com.

⁴ Searchable on public domain, all of Enron’s case data is available through eLaw Repository at <http://www.elaw4enron.com/default.asp>

⁵ Show Similar” is a powerful and unique Inference feature that allows users to instantly search the entire dataset for documents that are similar by concept.