

Maps, Clusters & Spikes: Exploring the Data Universe with Visual Analytics

By: Kevin Carr, President, InterLegis Inc.

A discovery database may be relegated to computers, but it can be as hard to embrace as a night sky filled with millions of stars. The enormity of it is daunting. It begins to make sense only when we're able to identify significant pieces ("stars"), connect them, and view them in small groups ("constellations").

This task need not be intimidating. We don't need a doctorate in astronomy to identify the salient objects in our solar system. Nor do we need an engineering degree to analyze a few million pages of discovered documents. We can do just fine with the deluge. We need only to visualize, meaningfully, what we've collected.

Visualization. It's a tool that truly can help us "conquer space"-data space-in the aftermath of the FRCP amendments.

Stars & Data Specks

In electronic discovery, we can break down our data universe; which typically is a virtual warehouse full of documents over which we're given custody in a pending case, into clusters of connected documents. Not only can modern conceptual analysis tools sort, organize and cluster these items; a cluster can be divided into subclusters. This approach makes it simpler for reviewers to grasp connections that will be crucial in developing case strategy. It simultaneously can help us quickly dispense with thousands upon thousands of irrelevant documents that muddle a data set. Bottom line: By clustering, we focus our time on the most meaningful documents and analyze them more effectively.

We can go further. Much like astronomers study starbursts in distant galaxies, data analysts can detect bursts and spikes in case activity (e-mail communication, for example) within a particular time frame.

In recent years, litigation support and e-discovery professionals have learned to use these and other advanced methods. Observing clusters and spikes can help us analyze and understand, easily and more quickly, the information we've amassed in discovery. This is essential when dealing with cases involving massive volumes of electronic data. We need to master search and review technology in order to successfully represent clients who are lost in data.

But these techniques don't go far enough. Electronic search results still confront us with a "flat list" of linear information. We need a way to quickly assess what we have.

Can You See It?

You've clustered your documents. You've identified spikes in activity. You've made connections-a lot of connections. And you're still overwhelmed.

The remedy is a newer form of technology called visual analytics. When employed early in the discovery process, visual analytics can help the litigation support team easily strategize and

prepare the case for trial or settlement negotiation. This technology can “picture” complex relationships among documents under review. It can show instantly such connections as the document author, creation date, file format and custodian. It also can graphically depict conceptual and meta data links. We can graph spikes and other significant activity trends, e-mail subject threads, and relationships among documents based on their unique attributes.

We also can “map” our data universe. Astronomers have created star maps, coordinating the locations of stars, galaxies and constellations. Navigators worldwide rely on them to chart courses and find their way from point to point. Modern search technology, likewise, renders maps of data that we have at our disposal. Mapping lets us cull our data collection and run ‘what-if’ scenarios before we begin full electronic data discovery processing. Data maps show search and review teams the overall picture.

When we thus visualize how our documents are interrelated, we can identify the critical ones much quicker. As we begin the document review, by employing visual analytics we literally can see the most important elements we have on our side (and notice which ones may be absent). By mapping out our document collection visually, we can see how the information relates within our discovery database and hone in on the results we need.

Trying New Combinations

With visual analytics, reviewers can test different combinations of information within the database to find special relationships and unique activity trends. For example, we can refine a list of documents to a particular date range, organize them by concept and then visually map them by custodian. Then we can see how these documents relate to other individuals by following communication threads. When we’ve identified the relationships, we can cluster the documents by file type and, within that group, reverse-cluster the documents back to date ranges.

Result: We’ve pinpointed a handful of documents that show clearly, by file type, specific communications that took place between parties during a given time period—and we’ve done it in a matter of minutes. Without visual analytics, the process could take weeks or months.

By employing visual analytics, electronic discovery professionals more easily can grasp the universe of information that has been turned over to them. Like stargazers locating constellations on a clear night, they gain an essential orientation for navigating through their data. They immediately can begin to organize documents into the most important sets and attack the review process far more effectively.

For overwhelmed litigation support professionals challenged with the complexity of electronic data discovery, the use of visual analytics just might make this task enjoyable again.

About the Author

Kevin Carr is president of InterLegis Inc., a Dallas-based company that works with corporate legal departments, law firms and legal service providers to reduce the risk, complexity and cost of litigation, regulatory requirements and internal investigations through early case assessment, electronic data culling and processing, hosted review, and advanced analytics. Kevin Carr can be reached at kcarr@interlegis.com.