Visual Analytics and e-Discovery: Seeing the Big Picture

By Victoria L. Lemieux
CiFER Research

For those who may be unfamiliar with the concept of Visual Analytics, it is not just another "pretty picture" - visualization tools by themselves are not enough to fully support someone who needs to examine and analyze large and complex sets of data. Rather, it is the science of analytical reasoning facilitated by interactive visual interfaces. It combines the best of human and machine intelligence to amplify cognitive processes. Some people liken it to "power tools for the brain." With true visual analytics tools analysts have access to artificial reasoning systems that can be integrated with their own reasoning processes. Analysts then have the ability to formulate hypotheses for a given situation, and discard data that doesn't apply as the program runs its analysis. The result is a more focused data set for analysts to examine and less time spent getting to that point.

For anyone familiar with the long, laborious and costly process of e-discovery, the advantage of technology that can help sift through huge datasets of diverse data types to identify what is discoverable from what is not is of huge benefit. With traditional tools of analysis, analysts face a long process of culling and de-duplicating large volumes of electronic datasets to produce relevant documents or emails, with associated metadata, stored in their native form, or in some normalized form such as EDRM XML, tiff, or pdf. The documents are then indexed and made available for search using standard search query types such as: term queries; phrase queries; near queries; range queries; wildcard queries; fuzzy queries and Boolean queries. The limitation of this approach is that the analyst may not be certain at the start of a case of the important topics, key players, important dates, or specific vocabulary in use. He may think he has an idea of what might be important and so try out a few simple queries to see what he can find. But his results may return zero finds. He may give up faced with the impossibility of searching for "unknown unknowns."

Visual analytics tools overcome this cognitive barrier by applying machine intelligence to analyse large datasets to "see the unseen." Since the human brain is able to understand information presented visually much faster than if the same information were to be presented in text form (as the old saying goes: "a picture is worth a thousand words"), visual analytics tools apply a number of "visual metaphors" such as node link diagrams, clusters or time lines to present the results of the machine analysis to the analyst. In true visual analytics tools (as opposed to tools that offer "visualization"), the analyst is then able to interact with the visual representations of the analysis to conduct further queries on the dataset. The results of these additional analytical steps also are presented as visual metaphors or representations of the data that can be further analysed using an iterative process.

In 2006, Forrester Research predicted that visual analytics was going to be "the next big thing" in e-discovery. The Forrester report said that "The largest addressable cost in eDiscovery is the cost of legal professionals who review data. Tools with visual analytics built in can make these legal professionals more efficient by determining whether or not data is relevant, is privileged, or even needs to be produced in response to a discovery request. Tools aimed at making lawyers with high hourly wages as efficient as possible represent the largest potential cost savings in eDiscovery tools. "[1] Since the report was issued, a number of technology vendors, both those that are e-Discovery "pure play" technology companies (e.g., MetaLINCS and Attenex) and those that offer more general search or data management software and services (e.g., Autonomy), have incorporated visualization into their e-discovery offerings. These tools range from those that offer visualization as an "add on" mode of displaying search results, which in itself can be extremely helpful to analysts, to those that offer "true" visual analytic capabilities - the ability to have a human analyst integrate with the visual analytics in order to really delve into data, to structure them, to organize them, and, ultimately, to understand them.

While visual analytics has penetrated the e-discovery technology market, it is far from a mature technology. The science of visual analytics (e.g., underlying algorithms and how these parse and represent data) is still evolving, the methods that analysts use to interact with the technology are not standardized, and the current e-discovery technology solutions only offer a fraction of the capabilities of more cutting edge visual analytics tools. Visual analytics as an emerging area of technology has great

[1] Barry Murphy, "Believe it – Ediscovery Technology Spending to Top $4.8 Billion by 2011," Forrester Research, December 11, 2006.

potential to improve efficiency and reduce cost in the e-discovery process. Its full potential has yet to be realised, however, and does not come without some risks. To find out more about what the potential benefits of deploying visual analytics in e-discovery are, what e-discovery analysis using visual analytics looks like, what the current state of development of visual analytics is as it applies to the field of e-discovery and where the future opportunities and pitfalls are in applying visual analytics to e-discovery, you are invited to attend the CiFER Research presentation on Visual Analytics at the Electronic Evidence and E-Discovery Forum, London, 20-21 October, 2009.

About CiFER Research

The Centre for the Investigation of Financial Electronic Records (CiFER) conducts research, hosts educational and networking events, and engages in advocacy related to records and information in financial institutions. We are funded by government grants, interest groups, businesses and from consulting and research work related to our mission. For more information about CiFER and to find out about our current projects and initiatives, go to www.ciferresearch.org