

The Metropolitan Corporate Counsel

www.metrocorpcounsel.com

Volume 18, No. 12

© 2010 The Metropolitan Corporate Counsel, Inc.

December 2010

Analytical Software: A Better Approach

The Editor interviews Conor R. Crowley, Esq., CIPP, an internationally recognized expert in e-discovery and e-compliance.

Editor: The cost of commercial litigation is ballooning as a result of dramatic increases in volumes of electronically stored information (ESI). Human review of all documents collected in a case is not feasible budget-wise or time-wise. What strategies are available to address this problem?

Crowley: One of the major changes we're seeing and will continue to see is a move away from human review because, as you indicated, it is not feasible time-wise or money-wise, and furthermore it does not allow litigants to comply with the mandate of Rule 1 of the Federal Rules of Civil Procedure, which is to resolve matters in a just, speedy and inexpensive fashion. I think we see a movement into the use of analytical software because it allows one to leverage the knowledge of experienced, knowledgeable counsel to make relevancy determinations that are demonstrably accurate without the need for individual attorneys to conduct a page-by-page review of an entire document collection.

Editor: Do you think the time will ever come when e-discovery will be totally automated?

Crowley: The days when software can operate completely independently are still a long way off because I think relevance and responsiveness are very subjective determinations that require the input of an attorney. For software to make that determination without human assistance is extremely difficult.

Editor: What is meant by use of the term "analytical software"?

Crowley: Analytical software is software that is used to make determinations as to the relevance of particular documents in a collection based on the content of the document rather than on the occurrence of specific words or phrases, i.e., a keyword search. For example, whereas a keyword search for "lawyer" or "plaintiff" would retrieve documents contain-

ing either of these words, analytical software could also retrieve documents involving discussions of lawsuits or defense arguments, but which do not contain the specific words "lawyer" and "plaintiff."

Editor: Is analytical software a monolith, or are there different approaches emerging?

Crowley: There are a variety of different offerings available that have slightly different approaches with each having its own set of advantages and disadvantages. Some simply gather similar documents into clusters, some make binary relevance determinations, and some, such as Equivio>Relevance, provide relevance rankings where they are actually rating the relevance of documents. This approach allows you to focus on the most relevant documents in a given matter.

Editor: Isn't the use of analytical software a relatively new approach to e-discovery?

Crowley: Analytical software itself is not new to the e-discovery space. For example, while not widely adopted, conceptual search and clustering have been present in the litigation market for a few years. New now is the emergence of more sophisticated techniques which use sampling and iterative learning. These approaches are new to e-discovery, although they are widely used in other applications, such as voice recognition or spam filters. These newly emerging approaches are based on machine-learning technology that improves its performance based on feedback from a human user. For instance, as the software is provided more information to assist in the determination of relevance with respect to a particular document population, its knowledge base and the accuracy of its determinations increase. This monitored, iterative approach enables greater accuracy, facilitating the defensibility of the process and opening the door for the use of this technology in filtering the document collection.

Editor: Please describe the advantages of this new generation of analytical software versus the typical binary keyword approach.

Crowley: One of the difficulties with keywords is that you need to have all variations of the spelling of a particular word, and you also need to track down antonyms and synonyms to ensure that you're retrieving all responsive documents. If you are not familiar with a particular lingo or jargon or you don't provide the proper term, you are not going to find the sought-after documents. With analytical software a lawyer reviewing documents and making determinations as to relevance based on the entirety of the document will give far more information to the software, allowing for retrieval of documents that don't necessarily contain specific keywords but which are relevant based on the concepts addressed and content contained in those documents.

Editor: Does analytical software sometimes also turn up documents that were not known to exist?

Crowley: It does because it is moving beyond the simple words and phrases and extrapolating from the information you give it to find documents that contain related content, even when those documents do not contain those words that you may have otherwise thought were the indicators of relevance. This allows for a higher level of recall – that is, a high rate of retrieval of relevant documents -- than is possible with simple keyword searches.

Editor: According to your recent article in the Bureau of National Affairs publication, keywords achieve an average recall rate of 22 percent. What is the corresponding success rate with analytical software?

Crowley: It varies based on the matter itself, the collection documents, and the software used. The beauty of the new iterative sampling tools is that you can continue to train the software until you achieve the levels of recall and precision that you're seeking. If you want to keep running iteration after iteration where you're providing more and more information to the software with respect to relevance and non-relevance, you can consistently get recall rates in the 80th percentile and above. One of the keys to remember: recall is only one of the measures of success! Precision is also impor-

Please email the interviewee at ccrowley@crowleylawoffice.com with questions about this interview.

tant because precision measures how many of the retrieved documents are actually relevant. This is important because it obviously impacts the time and money spent reviewing non-relevant documents.

Editor: What are the limitations of analytical software?

Crowley: Setting aside the differences between the various analytical software tools available, I think the limitations really are dependent upon the knowledge and experience of the human attorney that is guiding the software and providing it with the information to make relevancy determinations.

Editor: Are the federal courts accepting of analytical software?

Crowley: There has not yet been a decision where the use of analytical software to identify relevant documents by one party was challenged by the other, and that use was then found reasonable by a court. This is because there has not yet been a court challenge to the use of analytical software, not because its use is not defensible. There have, however, been a number of decisions by, for example, Judge Andrew Peck, Judge John Facciola and Judge Paul Grimm that acknowledge the advantages of analytical software or concept searching versus keyword searching. Also, the Advisory Committee notes to Federal Rule of Evidence 502(b), endorse the use of analytical software. FRE 502(b) provides that attorney-client privilege is not waived through the inadvertent disclosure of privileged materials provided the holder of the privilege took reasonable steps to prevent the disclosure. The Advisory Committee notes explicitly acknowledge that the use of analytical software to identify privileged materials may constitute "reasonable steps." With the volume of information that must be reviewed in complex litigation there really is no longer any feasible alternative to the use of analytical software. I think it's only a matter of time until there is a case where somebody challenges the use of analytical software. Insofar as the challenge relates to learning software that uses monitored, iterative sampling techniques, I am of the view that the court will deem such approaches reasonable and defensible for use in making relevance determinations in civil discovery.

Editor: What are the risks in using this kind of software?

Crowley: Most likely one of the grounds for challenge will be that human beings did not

review the documents because there remains a lingering belief that human review is the gold standard. Yet studies have shown that humans really aren't that great at finding relevant documents. The other problem with human review is that in a large case, first pass review can require such a large team that there is a strong likelihood of receiving inconsistent relevance determinations across the document population. The acceptance of keyword search, followed by high-volume human review, as the incumbent default process for e-discovery needs to be changed and challenged. If you have analytical software that has appropriate quality assurance where you're tracking the levels of recall and precision and where you're sampling documents that are deemed not relevant to ensure that relevant documents were not missed, I think the use of analytical software to refine the review set is more defensible and more reasonable than keyword searching or end-to-end human review.

Editor: In your opinion, is analytical software defensible?

Crowley: It is defensible with the qualification that its defensibility is very dependent on the process used. You need a process that employs quality assurance tools, that allows for sampling to ensure the determinations made are accurate, and that permits one to be able to give a report on the levels of precision and recall achieved by the software to show that the use of it was reasonable and satisfies one's obligations under the Federal Rules of Civil Procedure. For example, Equivio offers a predictive coding tool, known as Equivio>Relevance, which uses an iterative and monitored training process, and which is designed to provide multiple levels of quality assurance. These elements are important components in ensuring the defensibility of the process.

Editor: How do you see analytical software being deployed? Do you see it replacing human review?

Crowley: Not entirely. Currently we're in a situation where it is useful to leverage human knowledge by using analytical software, but where the software is not capable of making the subjective relevance determinations required without human input. I think you will see more and more use of it in the early case assessment context to identify relevant documents as quickly as possible, thereby allowing corporations to quickly assess the strengths and weaknesses of their cases. You are most likely to see its use in symmetrical litigation

where you have two corporate litigants who can agree that they will both use the same software to achieve agreed upon levels of precision and recall because they understand that this is a way to save tremendous amounts of money on discovery. In asymmetrical litigation there is unfortunately far less incentive for the party with very few documents to agree to the use of analytical software by the party with a large volume of documents to review. However, there is a valid argument to be made that not allowing the use of analytical software could result in a disproportionate burden on the party with the larger volume of documents to review, and the Federal Rules of Civil Procedure provide mechanisms to protect parties from discovery burdens that are not proportionate in a given matter.

Editor: How can privileged information be protected?

Crowley: That is really a very thorny problem because you have a situation where there is a belief that if one doesn't manually review documents for privilege one isn't fulfilling one's ethical obligation. But if you look at Federal Rule of Evidence 502, which was put into place to ensure the protection of attorney-client privilege and work product, it recognizes that the use of analytical software is appropriate to make these determinations. As with the use of analytical software to review for relevance, the use of analytical software to review for privilege will take a long time to completely replace human review but it can already be used today to assist lawyers in the identification of privileged materials.

Conor R. Crowley, a Certified Information Privacy Professional, is the principal of the Crowley Law Office. Mr. Crowley's practice advises corporations and law firms about best practices in e-discovery, e-compliance and data privacy, in addition to providing expert witness services. Mr. Crowley is a member of the Steering Committee for The Sedona Conference Working Group on Best Practices for Electronic Document Retention and Production, the Advisory Board for Georgetown University Law Center's Advanced E-Discovery Institute, and the Board of Advisors for BNA's Digital Discovery & e-Evidence. Mr. Crowley is also the Editor-in-Chief of the recently published Sedona Conference Commentary on Proportionality in E-Discovery, and a Senior Editor of The Sedona Conference Commentary on Legal Holds and The Sedona Principles (Second Edition): Best Practices Recommendations & Principles for Addressing Electronic Document Production.