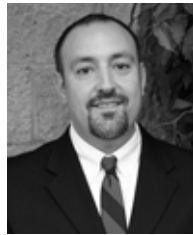

Reproduced with permission from Digital Discovery & e-Evidence, 11 DDEE 429, 10/27/2011. Copyright © 2011 by The Bureau of National Affairs, Inc. (800-372-1033) <http://www.bna.com>

BNA INSIGHT

Even with the advances in time and cost savings that eDiscovery technologies have provided, the issue of efficient and defensible privilege review remains complex and problematic. The authors take a comprehensive look at the issues and conclude that Predictive Coding is the optimal search technique in eDiscovery document review.

Predictive Coding Primer



BY ROBERT ALAN EISENBERG, ANNE S. PETERSON,
AND DANIEL D'ANGELO

From cellular phones, to personal computers and tablets, to mobile computing devices smaller than a deck of cards, technology has enabled people to connect instantaneously to an unending stream of data. These new technologies have allowed individuals and companies to stay connected and network in ways never before imagined while also providing the business world with the means to conduct transactions in an entirely digital fashion.

As we know, this onslaught of technology has drastically changed the legal world as well, especially as it pertains to trial practice. The outcome of litigation frequently turns upon a party's ability to discover or comprehensively and defensibly produce electronically stored information (ESI).

In this legal context, the enormous volume of ESI created and held in storage upon a wide universe of media has resulted in potentially massive amounts of data in need of investigation in order to determine if it is relevant to a particular matter. More importantly, the in-

formation must be reviewed in order to determine if the content is privileged or otherwise constitutes protected material.

Issues Within eDiscovery Review

It is this document review aspect of the discovery process that has, arguably, become the most procedurally challenging in terms of time and management. In addition, privilege review exposes attorneys to sanctions and/or malpractice liability in the event of inadvertent disclosure.

The potential for attorney mistakes resulting in severe penalties is arguably as likely an outcome during a complex privilege review as it is during the data preservation and litigation hold phases. Understandably, most lawyers view such reviews as the most costly and, in general, most daunting aspect of the discipline of eDiscovery.

Even with the advances in time and cost savings that eDiscovery technologies have provided, the issue of efficient and defensible privilege review remains complex and problematic. Experts estimate that conducting an

eDiscovery review can cost upwards of \$30,000 per gigabyte¹. This includes time for lawyers, consultants, and eDiscovery service providers to gather and sift through data as well as conduct the review process.

For an organization generating thousands of e-mails daily and storing information well into the terabytes, these costs can be staggering. They can be prohibitive and impact the issue of whether or not litigation should proceed in the first instance, despite the merits of the case.²

A significant issue generated by massive data stores subject to discovery centers upon the fear of relying primarily upon technology, especially “advanced technology,” in order to conduct automated reviews. Concerns regarding the defensibility of technology lead to hesitancy in adopting new systems for fear of malpractice claims, sanctions, and the waiver of privileged information that is inadvertently disclosed.

Dearth of Case Law. While there is nothing in the law that suggests that emerging technologies cannot be used in the place of more widely accepted automated processes and human review, case law does not presently exist that provides lawyers with definitive guidance as to what may or may not be permitted.

In a recent decision, however, the court quoted from the Federal Rules of Evidence Advisory Committee’s notes, which stated that: “. . . [D]epending on the circumstances, a party that uses advanced analytical software applications and linguistic tools in screening for privilege and work product may be found to have taken ‘reasonable steps’ to prevent inadvertent disclosure.”³ Nevertheless, what renders this situation even more problematic and the road to the adoption of new and more efficient technologies obstacle-strewn is that the legal community is generally not disposed, for mostly sound reasons, to exercise heroic or trailblazing efforts on the technological front.

The Role of Privilege Review In the Discovery Process

At the most elementary level, a privilege review requires careful examination of documents to determine whether or not they contain privileged or otherwise protected information (such as proprietary or highly confidential information) that is shielded and should not be routinely disclosed to an opposing party.

Scope. The review obligation is rendered daunting since the scope of discovery as defined by section 26(b)(1) of the Federal Rules of Civil Procedure (FRCP) includes any non-privileged information that is relevant to the subject matter of the action, whether or not the information is admissible at trial. The scope is potentially huge, as the discovery of any information that is reasonably calculated to lead to the discovery of admissible evidence is permitted⁴.

¹ David Degnan, *Accounting for the Costs of Electronic Discovery*, 12 Minn. Journal of Law, Science & Technology 1 (2011).

² See e.g., *Rodriguez-Torres v. Government Development Bank of Puerto Rico*, 265 F.R.D. 40 (2010).

³ *Datel Holdings LTD v. Microsoft*, No. C-0905535EDL (N.D. Ca. 2011)

⁴ Fed. R. Civ. P. 26(b)(1).

The major exception concerns privileged information. Even though privileged information is protected, the inadvertent disclosure of privileged information can act as a waiver of privilege under the Federal Rules of Evidence (FRE) and the use and subsequent disclosure of such information by an opposing party can be permitted under such circumstances.

This adds to the overall time expended and costs incurred in connection with review since counsel must ensure that review is conducted in a manner that reasonably assures that privileged information is not overlooked and inadvertently disclosed.

Help from Federal Rules. Portions of the FRCP and the FRE have been crafted to provide guidance to litigators in acting to prevent inadvertent disclosure.

Adherence to the “eDiscovery Rules” found in the FRCP has been relied upon to reduce the likelihood of inadvertent disclosure and the potentially disastrous results of such errant, however unintentional, an act. Rules that explicitly and implicitly address this danger are: FRCP 26(f)(3)(D) and 16(b) and the Meet-and-Confer requirement and other, both mandated and recommended, collaborative processes found therein and within the comments of the Rules Committee; FRCP 26(g)(1)(A) and (B) and its solidly grounded concept of “reasonableness” in the discovery of ESI; and FRCP 26(b)(5) with its explicitly stated process for the protection and return of inadvertently produced material that is claimed to be privileged or otherwise protected.

A *raison d’être* for these rules was to mitigate or eliminate the occasionally severe and sometimes even existential damage incurred by the inadvertent disclosure of protected evidentiary material, while moderating the cost of discovery. However, these rules have never been as effective as had been anticipated in providing such protection.

Enter FRE 502(b). With the adoption of FRE 502(b), more targeted and efficacious remedies were brought to bear in protecting against the “Sword of Damocles” of inadvertent disclosure dangling above the heads of counsel and client.

Waiver of Privilege and FRE 502(b)

Two events were significant in the creation of Rule 502(b). First, in order to prevent inadvertent disclosure and the waiver of privilege, litigants were sometimes expending vast amounts of effort and cost on pre-production privilege review or, at best, incurring costs out of proportion to the value of the underlying litigation.

Second was the decision by Chief Magistrate Paul Grimm of the United States District Court, District of Maryland, in the seminal case of *Hopson v. The Mayor and City Council of Baltimore*⁵, addressing both the concept of waiver and non-waiver agreements.

Hopson. The court in *Hopson* held that protective orders entered into by the parties prior to trial and deemed reasonable could provide protection from the waiver of privilege for documents that were inadvertently disclosed and deemed to be privileged.

The holding also provided that although such agreements were in place, they would not obviate the need of

⁵ *Hopson v. The Mayor and City Council of Baltimore*, 232 F.R.D. 228 (D. Md. 2005)

parties to undertake reasonable pre-production efforts in order to avoid inadvertent disclosure in the first instance, when doing so would not be unduly burdensome or expensive⁶.

Rule 502(b) provided that “disclosure of a communication or information covered by the attorney client privilege or work product protection . . . does not operate as a waiver in a Federal or State proceeding if (1) the disclosure is inadvertent; (2) the holder of the privilege or protection took reasonable steps to prevent disclosure; and (3) the holder promptly took reasonable steps to rectify the error, including (if applicable) following Federal Rule of Civil Procedure 26(b)(5)(B).”⁷

The crux of the new Rule is that inadvertent disclosure does not operate as a waiver if the holder took reasonable steps to prevent such a disclosure and employed reasonably prompt methods to retrieve the mistakenly disclosed information.

Victor Stanley. Although decided prior to the adoption of Rule 502(b), the holding in *Victor Stanley Inc. v. Creative Pipe Inc.*, provides valuable guidance in determining what constitutes an unreasonable privilege review.

The holding sets out circumstances which can result in waiver and the standard of reasonableness that must be adhered to in order to avoid such an outcome. In determining the “reasonableness” of the defendant’s actions to avoid waiver, the court looked at five balancing factors: (1) the reasonableness of the precautions taken to prevent inadvertent disclosure; (2) the number of inadvertent disclosures; (3) the extent of the disclosures; (4) any delay in measures taken to rectify the disclosure; and (5) the overriding interests in justice⁸.

The balancing test used in *Victor Stanley* stresses the importance that privilege review holds and that counsel must adhere to reasonable standards that will be weighed by the courts in determining if waiver has occurred.

Reasonableness and Privilege Review

A major theme in both FRCP 26(g) (1) and FRE 502(b) is the concept of reliance upon a reasonable approach to prevent inadvertent disclosure and the waiver of privilege. But what is *reasonable*? Is the use of highly sophisticated, technology to conduct privilege review reasonable? Case law established the precedent that for a search to be deemed adequate, it is judged by a reasonableness standard⁹. But, again, what is the definition of a reasonable standard?

Unfortunately, there exist no clear answers in the case law or in commentaries. And lawyers, by nature, fear such uncertainties. To date, most sophisticated forms of technology in the review process center around the use of “advanced search” technologies. However, even with a comfort level building in connection with the use of generally familiar advanced search technologies, questions still arise as to its reasonableness and the capability of attorneys and courts to make

the determination that the reasonableness standard has been attained.

The Rise of Predictive Coding

In order to achieve the optimum results of reduced time and costs in performing privilege review utilizing a reasonable and defensible advanced search technique, the legal community undertook the process of evaluating new technologies. On the forefront of that inquiry is the review system known as Predictive Coding.

And, indeed, while the jury is still out on the subject, arguably, Predictive Coding is the optimal search technique in eDiscovery document review.

What is Predictive Coding? Predictive Coding is certainly not new, nor exclusive to the legal world. Indeed, Predictive Coding is considered more of a process that blends human skill and technology rather than a singular technology.

Predictive Coding, in general, has been defined as “the electronic . . . organization, and prioritization of entire sets of ESI according to their relation to discovery responsiveness, privilege, and designated issues before and during the legal discovery process.”¹⁰

At a very high level, Predictive Coding involves using a skilled attorney to seed a Predictive Coding engine with a small set of documents that are selected based on their relevancy or privileged content. The Predictive Coding software then analyzes the information contained in the initial set of seed documents to identify references in the text such as people, concepts, or places to formulate rules that then shape the search for information of that type throughout the universe of documents¹¹.

As with most sophisticated document review technical procedures, this process is iterative, and results that are returned can be refined and subsequent passes by the software’s algorithms can find more precise matches in the document collection. As results are refined, the criteria used in the search can then be “propagated” or passed along to the entire collection of potentially discoverable documents.

Relation to Concept Search. Predictive Coding is grounded upon the science of “Concept Search.” In general, Concept Search works by modifying the user’s inquiry to take advantage of the fact that most words can represent more than what they signify merely standing alone.

For instance, in using a concept search-based technology, a search for “car” will automatically be modified to a search for words conceptually related to car, such as “truck” or “vehicle.” Or a search for the word “Arctic” will programmatically return related words, such as “Eskimo,” “igloo,” and “polar bear.”

This process of modifying an inquiry beyond the single term used is known, in the world of Concept Search, as “Query Expansion” and all types of Concept

⁶ Id.

⁷ F.R.E. 502

⁸ *Victor Stanley Inc. v. Creative Pipe Inc.*, 250 F.R.D. 251 (D. Md. 2005)

⁹ *Security Financial Life Insurance Company v. Department of Treasury*, 2005 WL 839543, 4 (D.D.C. 2005).

¹⁰ Ben Kerschberg, *eDiscovery and the Rise of Predictive Coding*, <http://blogs.forbes.com/benkerschberg/2011/03/23/eDiscovery-and-the-rise-of-predictive-coding/> (March 23, 2011).

¹¹ Caitlin Murphy, *5 Things You Should Know About Predictive Coding*, <http://ediscoveryinsight.com/5-things-you-should-know-about-predictive-coding/> (January 25, 2011).

Search technologies, including Predictive Coding, involve such Query Expansion.

Concept search can be used to identify responsive documents even when the correct words to search for are unknown. The concept search will display at the top of the search results those documents that are most relevant.

Moreover, the search will enable the inquirer to understand the manner in which words were used in the document and demonstrate the conceptual relationship between the search words and their meaning.

eDiscovery Institute's Survey On Predictive Coding

A recent survey conducted by the eDiscovery Institute examined Predictive Coding and its acceptance in the eDiscovery field¹². The Predictive Coding survey was conducted with 11 eDiscovery service provider participants employing some form of Predictive Coding in their technology.

Notably, the most telling finding of the survey indicated that the use of Predictive Coding systems represented significant cost savings compared to traditional review. Reported results showed an average savings of 45 percent with an average maximum of 71 percent and average minimum of 23 percent. Individual respondents reported cost saving highs of 80 percent, 95 percent, and 100 percent¹³.

Another significant finding was that nine of the 11 respondents found that their technology produced the same repeatable results in direct contrast to traditional human review.

Aside from cost savings and repeatable results, the survey also highlighted other inherent advantages to using Predictive Coding. Because the costs of linear review are typically very high, counsel does not have the ability to reevaluate data sets. Predictive Coding aids reviewers by allowing sets of data to be more economically examined multiple times, in the course of which, the results are strengthened with each iteration.

Predictive Coding solutions also provide transparency and audit trails that reveal what decisions were made in the review process; a potentially powerful response to reasonableness questions that may arise.

These solutions also shortened the time frame of a traditional review as larger volumes of data could be reviewed more rapidly.

Predictive Coding Issues And Obstacles to Adoption

Lack of Awareness. While highlighting the strong points of Predictive Coding, the survey also emphasized the issues plaguing wider use and adoption of the technology. Certainly, lack of awareness in the legal community is a contributing factor in constraining the broader acceptance of the use of Predictive Coding. This may result partly from attorneys not understanding the technology and partly from reluctance in becoming a pioneer in the use of any process that departs

from standard or more commonplace technical solutions.

Uncertain Judicial Reception. However, more specifically, the most cited reason that Predictive Coding has not gained widespread use is the uncertainty and fear concerning whether courts would accept Predictive Coding under the reasonableness standards of Rule 26(g)(1) and Rule 502 (b). The possibility of counsel using Predictive Coding to conduct a privilege review that the court then deems unreasonable, resulting in a waiver of privilege, is a risk that many practitioners are not willing to take.

As with most technological solutions in any field of endeavor, such trepidation will be reduced as the approach is more frequently utilized and encountered by counsel, client, and the bench. It is likely that the cost of review, especially in these economically fraught times, will, inevitably drive the use of Predictive Coding (PC) forward, until PC becomes, in practice, indeed, more "PC".

What We Fear. Herb Roitblat, one of the pioneers in the Advanced Search field, has succinctly described the main conundrum concerning the science of conceptual searching, of which, as mentioned, Predictive Coding is considered a sophisticated weapon within the document review arsenal. In reference to concept search, including Predictive Coding, Dr. Roitblat, who grapples determinably with allaying fear daily, has noted that attorneys (soft science acolytes that they are), are frequently driven by fear such as:

- Fear of not knowing how to use the tool;
- Fear of not knowing how to explain the use of the tool;
- Fear of not knowing how to evaluate the tool's efficacy;
- Fear that counsel is being deliberately misled by service providers promoting the tool.

Although, in general, an all too accurate statement of many practitioners' present attitude toward advanced review technology in eDiscovery Practice, as Roitblat has pointed out, these fears will become manageable, as the efficacy and, consequently, the defensibility, of Predictive Coding is validated with its increasingly frequent use.

The Defensibility of Predictive Coding

Familiarity Breeds Acceptance. Accordingly, the highest hurdle to the more wide-spread use of Predictive Coding is firmly establishing its defensibility in conducting privilege review. Nevertheless, privilege review conducted using Predictive Coding will be upheld if it comports with FRCP 26(g)(1) and FRE 502 (b).

As mentioned, Rule 26(g)(1) requires an attorney to attest that a discovery response was conducted after a reasonable inquiry was made¹⁴, and, Rule 502(b) protects a waiver when disclosure was made inadvertently and the holder of the privilege took reasonable steps to prevent its disclosure¹⁵.

¹² Anne Kershaw & Joe Howie, *eDiscovery Institute Survey on Predictive Coding* (eDiscovery Institute 2010).

¹³ Kershaw & Howie, *supra* n. 33, at 3.

¹⁴ Fed. R. Civ. P. 26(g) (1).

¹⁵ F.R.E. 502(b).

In sum, in order for Predictive Coding to be more widely embraced and used in privilege review, its use must be considered reasonable under these Rules and the trend indicates that with the growing familiarity of bench and bar it most likely will be.

Presently, many writings on the subject and almost as many eDiscovery vendors claim that Predictive Coding is, in fact, imminently, defensible; however, those claims have little or no significance to a court. Unfortunately, case law does not exist that is directly on point concerning the use of Predictive Coding in conducting privilege review.

The Victor Stanley Factors. For any review, the route to the “Holy Grail” of Predictive Coding acceptability lies in the five factor test presented in Judge Grimm’s, *Victor Stanley* decision, which remains the penultimate guidance for the manner in which reasonableness will be judged¹⁶.

It is also noteworthy that although opposing counsel may object to the use of Predictive Coding as an insufficiently proven, overly-complex, and, accordingly, unreasonable and ineffective search methodology, a review of the case law will bear out that most courts do not typically support objections based on the mere supposition that some evidence may have been missed. The opposing party must respond with some particularity and that requirement may be an insurmountable obstacle to an effective assault on reliance upon Predictive Coding in many instances¹⁷.

Bottom Line on Predictive Coding as Reasonable Privilege Review Tool

Although no case law exists that clearly defines reasonableness in the use of Predictive Coding, newly published research supports the efficacy of this technology by documenting the outcome of the pitting of man against machine. The results may help to alleviate some of the concerns regarding whether new and innovative concept search technology can be considered reasonable.

In her article, *Automated Document Review Proves Its Reliability*, Anne Kershaw described a study that she conducted in which a set of 48,000 documents were reviewed and coded for relevance in three responsive categories. The study tested software with the capability to review and tag the documents against six experienced human reviewers. The software and reviewers separately reviewed the documents.

Across all three responsive coding passes, the software identified more than 95 percent of relevant documents with a high of 98.8 percent for one pass. In comparison, the human reviewers averaged only 51.1 percent with a low of 43 percent on one of the coding iterations.¹⁸

In a more recent study, human reviewers were again compared with an automated review system, and responsiveness rates proved nearly identical. “The results support the concept that machine categorization is, at

¹⁶ See e.g., *Amobi v. District of Columbia Dept. of Corrections*, 262 F.R.D. 45 (D.D.C. 2009), *Mt. Hawley Ins. Co. v. Felman Production, Inc.*, 271 F.R.D. 125 (S.D.W.Va. 2010).

¹⁷ *Ford Motor Co, et. al, v. Edgewood Properties, Inc.*, Civil Action No. 06-1278 (HAA) (ES) (D.N.J. 2009)

¹⁸ Kershaw, *supra* n. 5, at 3.

least, no less accurate at identifying relevant/responsive documents than employing a team of reviewers. Based on these results, it would appear that using machine categorization can be a reasonable substitute for human review¹⁹.”

In objectively analyzing the results of these studies, even a non-expert will concur that they make practical sense in the reduction of cost and risk exposure, since a computer can perform the same task repeatedly without exhaustion or distraction. The conclusion, then, is that in most instances, technology can perform a review that provides better results than human review while using less manpower, consuming less time, and costing less money. Accordingly, the use of document review technology, including Predictive Coding, would seem defensible and soundly reasonable.

The Future of Predictive Coding

These studies, and several others like them, demonstrate that this is an area where the legal and technological communities can work together to achieve an optimum end result: advanced technology that can be used to conduct a truly economic privilege review that will be deemed reasonable and be completely defensible in any court.

A driving force behind this push is The Sedona Conference®. To illustrate the collaborative effort needed to advance review technology, The Sedona Conference® has been partnering with The Text REtrieval Conference (TREC). Its purpose is to support research within the information retrieval community by providing the infrastructure necessary for large-scale evaluation of text retrieval methodologies²⁰. TREC has a legal track devoted to developing search technology that meets the needs of lawyers to engage in the effective discovery and review of ESI²¹.

Blending experts in the legal arena with the technology world ensures that a solution is viable. While a viable solution, in part, revolves around the accuracy and reliability of the software to perform a privilege review, the human element in a privilege review for privilege cannot be diminished. As noted, in Predictive Coding, attorneys must first seed the solution in order to provide the proper search parameters and must then review the results. With that in mind, it is crucial that technologically savvy attorneys continue to work with technical solution providers to help mold and soundly validate the worth of Predictive Coding.

Final Thoughts

The effort to gain general acceptance of Predictive Coding and other advanced search document review technology is gaining momentum. In January of this year, an eDiscovery service provider helped analyze 1.5 million documents for less than \$100,000. Similarly, an-

¹⁹ Herbert L. Roitblat, Anne Kershaw, Patrick Oot, *Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review*, 61 Journal of the American Society for Information Science and Technology 1 (October 8, 2009).

²⁰ Text REtrieval Conference (TREC), *Overview*, <http://trec.nist.gov/overview.html> (accessed May 5, 2011).

²¹ Text REtrieval Conference (TREC), *Tracks*, <http://trec.nist.gov/tracks.html> (accessed May 5, 2011).

other eDiscovery vendor provided software that enabled a law firm to search through a half-million documents under a draconian court-imposed deadline of one week. The software analyzed and sorted 570,000 documents, some having multiple pages, in two days. The law firm consumed just one more day to identify 3,070 documents that were relevant to the court-ordered discovery motion²².

It would seem undisputable that technology can benefit attorney and client, since they are enabled to conduct discovery in a more effective and more cost efficient manner, and, in turn, reduce the overall time in which the litigation is conducted.

More important are the benefits of Predictive Coding in safeguarding privilege and preventing the inadvertent disclosure of information that has the potential of determining the outcome of litigation upon issues other than its merits. It also seems a reasonable assumption, that, in fact, the apprehension in the use of Predictive Coding for privilege review does not arise from a reasoned evaluation of the technological solution itself; but rather it is embedded in the conditioned mindset of the culturally resistant legal community that must embrace it. To this cultural mindset one should add as a barrier to the acceptance of Predictive Coding the economic interest law firms have in maximizing fees realized by privilege review.

Nevertheless, it would appear that widespread adoption is within reach. But, as with any significant advance in whatever human endeavor, technologically based or otherwise, someone has to “go first” in demonstratively challenging the conventional wisdom with that one very prominent, high-profile, exemplar scenario. We still await such a matter.

²² John Markoff, *Armies of Expensive Lawyers, Replaced by Cheaper Software*, N.Y. Times A1 (March 5, 2011).

Robert Alan Eisenberg is the Director of eDiscovery Services for OrcaTec LLC, based in New York City and Washington DC. Mr. Eisenberg has over 25 years experience in the practice of law and the rendering of consultative services relating to the discovery of Electronically Stored Information (ESI), computer forensics, proactive electronic data preservation and retention and traditional paper-based litigation support services. He is the founder and Permanent Co-Chair of The Advanced eDiscovery Institute and the eDiscovery Training Academy at the Georgetown University Law Center in Washington DC. Mr. Eisenberg is also the Program Chair of Georgetown Law's eDiscovery Practice Support Project and serves upon the Advisory Board of BNA's Digital Discovery & e-Evidence (DDEE). He can be reached at robert.eisenberg@orcatec.com

Anne S. Peterson is the Director of Discovery Compliance and Information Management at Precise Inc. Ms. Peterson also serves as the Director of E-discovery Education and Services at Duquesne University School of Law, where she teaches the first e-discovery course in the country to use live search and review software in a law school environment. She developed and copyrighted the course which uses a case study and “hands on” approach with the goal of graduating lawyers who are technically competent in e-discovery.

Daniel D'Angelo is a 2011 graduate of Duquesne University Law School where he graduated cum laude. Mr. D'Angelo is also a Project Program Management Advisor with Dell, Inc.; he has over 10 years of experience working as a project manager and business analyst designing and implementing business solutions based on Microsoft applications. Mr. D'Angelo passed the July 2011 Pennsylvania Bar and is in the process of completing his admission to practice in the Commonwealth.