

Skadden

Beijing
Boston
Brussels
Chicago
Frankfurt
Hong Kong
Houston
London
Los Angeles
Moscow
Munich
New York
Palo Alto
Paris
São Paulo
Shanghai
Singapore
Sydney
Tokyo
Toronto
Vienna
Washington, D.C.
Wilmington

ESI and Predictive Coding

Charles W. Schwartz

Chris Wycliff

December 13, 2013

In the Beginning

“...By computer assisted coding, I mean tools (different vendors use different names) that use sophisticated algorithms to enable the computer to determine relevance, based on interaction with (i.e., training by) a human reviewer unlike manual review, where the review is done by the most junior staff, computer-assisted coding involves a senior partner (or team) who can review and code a “seed set” of documents. The computer identifies properties of those documents that it uses to code other documents. As the senior reviewer continues to code more sample documents, the computer predicts the reviewer’s coding. (Or, the computer codes some documents and asks the senior reviewer for feedback.)”

Judge Andrew Peck “Search Forward” Law Technology News October 1, 2011

What is Predictive Coding?

- Traditionally, electronically stored information (ESI) has been culled down by a combination of keyword searches, filters (e.g., date and custodians) and human review.
- Predictive coding uses computer programs and algorithms to determine whether documents are relevant to a case.
- Predictive Coding technology is similar to email spam filters which train software to identify spam.

How does Predictive Coding Work? - 1

- Predictive Coding works by training a computer program to identify responsive documents.
 - There is more than one way to select a source set of potentially relevant documents. This is how the workflow for the predictive coding paradigm commences. Below are three of the most popular methodologies:
 - Running the software against the entire data collection where the software makes the first initial cut of relevancy
 - Narrowing the entire data collection by using search terms
 - Senior attorneys, who are familiar with the details of the case, select a limited batch of relevant, non-privileged documents

How does Predictive Coding Work? - 2

- Using these methodologies, the computer program attempts to identify the truly responsive documents within the source set.
- Usually multiple iterations are necessary to achieve a satisfactory level of confidence that the program is identifying the desired percentage of responsive documents. This is part of the training process.
- As the program learns to distinguish what is relevant, each iteration produces a smaller relevant subset of documents and a larger set of irrelevant documents (the null set) that can be used to verify the integrity of the results by confirming the absence of any probative material therein.

Concerns

- No real case law standards in place.
- Increased involvement from opposing counsel, such as in reviewing producing party's documents, e.g. opposing party may be involved in determining what documents are relevant, which is the domain of the producing party.

Predictive Coding

- Faster, but
 - Time to Consult/Coordinate/Negotiate with Vendor, Adversary, Court, Special Master
- Not Necessarily Less Expensive
- Strategy (Seed Sets – The New Key Words)
- Machines are more Consistent than Humans
- Types of ESI – Works Best with Text-Based Documents (not Excel Spreadsheets)

Predictive Coding

- Defensibility
- Transparency
- Ability to:
 - Prioritize the Review
 - Relevancy Ranking
 - Eliminate Irrelevant Documents
 - Test and Validate Human Coding Decisions
 - Identify More Responsive Documents
 - Supplement Key Word Lists
 - Categorize and Rank Third Party Productions

Predictive Coding Workflow

Validation and Production

- Validate Results
 - Statistical Quality Control
 - Random Sampling
- Review of Non-Responsive ESI
 - Testing and Sampling
- Review of Documents
- Privilege Logs, Redactions, Confidentiality

- Case Law

Current ESI Discovery Practices in the Courts

- ***Da Silva Moore v. Publicis Groupe & MSL Group***
(S.D.N.Y. February 24, 2012)
 - First opinion in which Court approved use of computer assisted review
 - Gender discrimination case; ~3M potentially responsive ESI docs
 - Both sides agreed to use predictive coding, but initially disagreed about the details
 - Opinion attaches the parties' detailed ESI review protocol
 - Court highlights importance of cooperation of the parties and direct involvement of experienced attorneys in the creation of the seed documents used to train and stabilize the review algorithm
 - Court praised Defendant for the transparency of its review process

Current ESI Discovery Practices in the Courts

- *Da Silva Moore (cont'd)*
 - Conclusions of United States Magistrate Judge Andrew J. Peck:
 - “[C]omputer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review.”
 - “Computer-assisted review now can be considered judicially-approved for use in appropriate cases.”
 - “[This] does not mean computer-assisted review must be used in all cases, or that the exact ESI protocol approved here will be appropriate in all future cases that utilize computer-assisted review.”
 - Ruling affirmed by District Court Judge Andrew L. Carter, Jr. (April 26, 2012)

Current ESI Discovery Practices in the Courts

- ***Global AeroSpace Inc. v. Ladow Aviation, L.P.***
(Loudon Cir. Ct. Va., April 23, 2012)
 - Design/construction defect case; ~ 2M potentially responsive electronic documents
 - Defendants sought to use predictive coding; Plaintiffs objected
 - Court ordered use of predictive coding over Plaintiffs’ objection because “the most effective and economical means of reviewing large ESI collections is a technology known as predictive coding.”
 - The order was entered “without prejudice to [plaintiffs to raise] completeness or the contents of the production or the ongoing use of predictive coding.”

Current ESI Discovery Practices in the Courts

- ***National Day Laborer Organizing Network v. United States Immigration and Customs Enforcement Agency***
(S.D.N.Y. July 13, 2012)
 - Challenge to reasonableness of FOIA search efforts
 - Judge Scheindlin comments re “best practices” for document collection and review:
 - “There is increasingly strong evidence that ‘[k]eyword search[ing] is not nearly as effective at identifying relevant information as many lawyers would like to believe.’”
 - “There are emerging best practices for dealing with these shortcomings There is a ‘need for careful thought, quality control, testing, and cooperation with opposing counsel in designing search terms. . . .’”
 - “[B]eyond the use of keyword search, parties can (and frequently should) rely on [computer assisted or predictive coding] tools to find responsive documents.”
-

Current ESI Discovery Practices in the Courts

- ***In Re: Actos*** (W.D. La. July 27, 2012)
 - Pharmaceutical MDL
 - Agreed-upon order entered that spells out detailed protocol
 - Both plaintiffs and defendants heavily involved in the algorithm “training” process
 - Extensive quality control via random sampling
 - Defendants may withhold documents for non-relevance, confidentiality, and privilege after predictive coding; Plaintiffs may audit withheld documents by statistical sampling
 - Protocol requires ongoing meet and confer sessions including custodian selection and post-production sampling to “finalize the method for searching documents on a going forward basis.”

Current ESI Discovery Practices in the Courts

- ***Kleen Products LLC v. Packaging Corporation of America*** (N.D. Ill. September 28, 2012)
 - Consumer product antitrust action; Defendant had produced in excess of 4M pages at time of order
 - Plaintiffs objected to Defendants’ use of Boolean searches instead of using predictive coding
 - Court noted Sedona Conference Principle 6: “[r]esponding parties are best situated to evaluate the procedures, methodologies, and techniques appropriate for preserving and producing their own [ESI].”
 - Court refused to order a re-review of documents using predictive coding; however, the parties agreed to meet and confer “regarding the appropriate search methodology to be used” for future document collections

Current ESI Discovery Practices in the Courts

- ***EORHB, Inc. v. HOA Holdings LLC***
(Del. Ch. Oct. 15, 2012)
 - Following oral argument on partial summary judgment and motion to dismiss, Vice Chancellor J. Travis Laster addressed the topic of a scheduling order and, *sua sponte*, ordered the use of predictive coding:
 - “This seems to me to be an ideal non-expedited case in which the parties would benefit from using predictive coding. I would like you all, if you do not want to use predictive coding, to show cause why this is not a case where predictive coding is the way to go.”
 - Parties ordered to meet and confer on the joint use of a single ESI vendor; Court would select a vendor if the parties could not agree

Current ESI Discovery Practices in the Courts

- ***In re Biomet*** (N.D. Ind. April 18, 2013)
 - Product liability MDL; document review and production began before formation of the MDL docket
 - Steering Committee objected to the pre-MDL ESI review process
 - The defensible results of the pre-MDL review process led to denial of Steering Committee’s request for re-review of data as disproportionate under Rule 26(b)(2)(C)
 - Court found Defendant met its discovery obligations when it used keyword searches and a de-duplication process to reduce custodian data from 19.5M to 2.5M documents before employing a predictive coding algorithm
 - Defendant had offered to let the Steering Committee choose additional search terms

Current ESI Discovery Practices in the Courts

- Key elements of predictive coding cases
 - Cooperation between the parties
 - Parties should meet and confer to choose keywords used to select or supplement the seed set used to “train” the algorithm
 - Parties should meet and confer regarding the results of the iterative review rounds
 - Production of the entire seed set including non-responsive documents
 - Defensible results
 - High recall and precision rates
 - Quality control testing
 - Transparency of process
 - Proportionality
 - Staged discovery beginning with the most likely relevant sources
 - Acknowledge that it is not possible to locate and produce every responsive piece of ESI