



Battling Information Inflation With Predictive Coding And Other Methods

Tom Myrick

Moore & Van Allen (Charlotte, NC)

tommyrick@mvalaw.com | 704.331.1126

<http://www.mvalaw.com/professionals-608.html>

BATTLING INFORMATION INFLATION

What It Is, How It Works, & Why
Should You Care

Thomas (Tom) D. Myrick

tommyrick@mvalaw.com

M|V|A
MOORE AND VAN ALLEN

OVERVIEW

- ❑ Discovery Technology Timeline
- ❑ Predictive Coding Defined
- ❑ Predictive Coding Workflow
- ❑ Key Benefits of Predictive Coding
- ❑ Other Technology Assisted Review ("TAR") Tools
- ❑ How Does the Judiciary View PC?
- ❑ What are the keys for defensibility?

DISCOVERY TECHNOLOGY TIMELINE

Early 1990s - 2000s

- Document Imaging
- OCR
- Text Extraction
- Key Term Searching



Mid 2000s

- Email explodes in the work place
- Email Threading
- Near Dedup Detection
- Web Based Review Tools



2013

- Social Media Discovery
- Cloud Apps and Collection
- Predictive Coding
- TAR



DISCOVERY TECHNOLOGY TIMELINE

Early 1990s - 2000s

- Document Imaging
- OCR
- Text Extraction
- Key Term Searching



□ Document Imaging

- Scanning hard copy paper documents into image formats for viewing within review software (Summation / Concordance)

□ Optical Character Recognition

- Computer software renders searchable text from scanned hard copy paper documents

□ Text Extraction

- Text within eDocs and email is accumulated into searchable text files

□ Key Term Searching

- Series of terms applied to the document universe to cull for privilege or responsiveness

DISCOVERY TECHNOLOGY

Early 1990s - 2000s

- Document Imaging
- OCR
- Text Extraction
- Key Term Searching



Mid 2000s

- Email explodes in the work place
- Email Threading
- Dedup Detection
- Web Based Review Tools



□ Email Revolution

- Email becomes ubiquitous within the workplace

□ Email Threading

- To help review teams, software used to group email discussions together

□ Dedup Detection

- MD5 Hash algorithm used to remove exact duplicate documents. "Near" dedup techniques developed to cull documents with subtle differences

□ Web Based Review Tools

- Thin client web based hosted review tools rise in popularity

DISCOVERY TECHNOLOGY TIMELINE

- ❑ **Social Media Discovery**
 - Discovery bleeds into social networks such as Twitter & Facebook
- ❑ **Cloud Apps and Collection**
 - Online tools such as Dropbox, Google Gdrive, Microsoft Cloud Drive are sources of documents subject to subpoena requests
- ❑ **Predictive Coding**
 - The explosion of large data sets and the soaring costs of document reviews, pushes the adoption of Predictive Coding and other TAR techniques
- ❑ **TAR**
 - Email Discussion Threading, Document Clustering, Near-Dedup detection,

2013

- Social Media Discovery
- Cloud Apps and Collection
- Predictive Coding
- TAR



PREDICTIVE CODING DEFINED

“[a]n industry specific term generally used to describe a [TAR] process involving the use of a machine learning algorithm to distinguish relevant from non-relevant documents, based on subject matter expert(s)’ coding of a training set of documents.”

-- Grossman-Cormac TAR Glossary

PREDICTIVE CODING DEFINED

- ❑ Type of machine-learning technology that enables a computer to help “predict” how documents should be classified based on limited human input
- ❑ The process entails automatically feeding decisions made by attorneys about the responsiveness of a small number of case documents called a *training set* into a computer system
- ❑ The computer relies on these training decisions to create a model that automatically generates a prediction score for every document based on the document’s degree of responsiveness

PREDICTIVE CODING DEFINED

- ❑ Uses a “active machine learning approach” to distinguish relevant from non-relevant documents, based on decision made by a subject matter expert (“SME”)
- ❑ Combines the efficiencies of a computer system with the knowledge of human beings
- ❑ Uses established statistical principles to measure status and accuracy

THE PROBLEM

- ❑ In the course of business, documents are not organized by relevance
- ❑ Clients tend to provide documents as they were stored



PREDICTIVE CODING

- ❑ SME reviewers train the software on relevance as the system samples the population

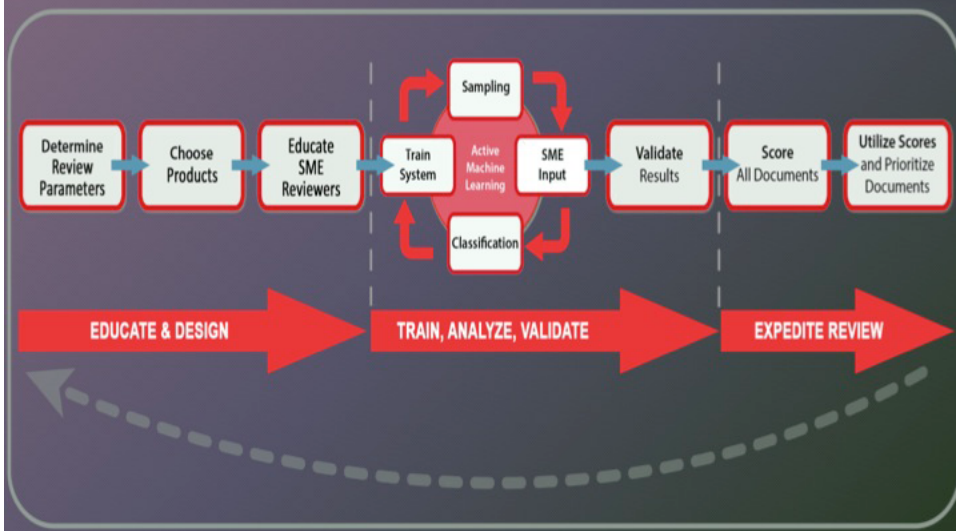


END RESULT

- ❑ The Predictive Coding application suggest documents that may be responsive, and the SME review team accepts or rejects the suggestions. As the SME review team reviews and codes the suggested documents based on the relevance to the case, the system learns more about the case.
- ❑ The predictive coding model is then applied to the remaining documents in the population.



PREDICTIVE CODING WORKFLOW



KEY BENEFITS

- ❑ Knowledge gained by manually reviewing a small percentage of the documents gets consistently applied to the entire universe of documents
- ❑ The training stage often reveals sufficient material for a better case assessment (first 2 to 4 days)
- ❑ Significant reduction in overall document review cost
- ❑ Significant reduction in overall time to complete a fully defensible document review
- ❑ Gaining acceptance by the judiciary

OTHER TAR TOOLS

- ❑ **DISCUSSION THREADING**
 - Utilizes algorithms to dynamically link together related documents (most commonly e-mail messages) into chronological threads that reveal entire discussions
 - This simplifies the process of identifying participants to a conversation and understanding the substance of the conversation
- ❑ **CLUSTERING**
 - Involves the use of algorithms that automatically organize a large collection of documents into different topical groupings based on similarity
- ❑ **NEAR-DUPLICATE IDENTIFICATION**
 - Allows SME reviewers to easily identify, view, and code near-duplicate emails, attachments, and loose files
 - Some systems can highlight differences between near-duplicate documents to help simplify document review

OTHER TAR TOOLS

❑ KEYWORD SEARCH

- Involves inputting a word or words into a computer which then retrieves documents within the collection containing the same words
- Also known as Boolean searching, keyword search tools typically include enhanced capabilities to identify word combinations and derivatives of root words among other things

❑ CONCEPT SEARCH

- Involves the use of algorithms to determine whether a document is responsive to a particular search query
- The technology typically analyzes variables such as the proximity and frequency of words as they appear in relationship to a keyword search

SOFTWARE PRODUCTS



HOW DOES THE JUDICIARY VIEW PREDICTIVE CODING?

- ❑ **Da Silva Moore v. Publicis Groupe**
 - Court okayed parties' agreement to use PC; parties disputed methodology (3.3 million emails)
- ❑ **Kleen Products v. Packaging Corp. of Am.**
 - Plaintiffs abandoned arguments in favor of PC and moved forward with Boolean search
- ❑ **Global Aerospace Inc. v. Landow Aviation, L.P.**
 - Court approved defendant's use of PC over plaintiff's objections (2 million emails)
- ❑ **Actos (Pioglitazone) Products Liability Litigation**
 - Court affirmatively approved the use of PC for review and production
- ❑ **EORHB, Inc., et al v. HOA Holdings, LLC**
 - Court orders parties to use PC and share common eDiscovery provider
- ❑ **In re Biomet M2a Magnum Hip Implant Prods.**
 - Initial use of key terms, then PC to cull document population upheld by court
- ❑ **FRE 502**

DEFENSIBLE PREDICTIVE CODING

- ❑ **Using Da Silva as a map**
 - Senior attorneys must be involved
 - Cooperate in devising approach
 - Have a written protocol
 - Share the "Seed Set"
 - Refine repeatedly for accuracy
 - Be transparent
- ❑ **Bottom line for defensibility**
 - Satisfy the rules and document your actions

FRE 502 ADVISORY COMMITTEE NOTES

- ❑ The Judicial Conference Rules Committees have long been concerned about the rising costs of litigation, much of which has been caused by the review, required under current law, of every document produced in discovery, in order to determine whether the document contains privileged information.

FRE 502 ADVISORY COMMITTEE NOTES

- ❑ Inadvertent disclosure of protected communications or information in connection with a federal proceeding or to a federal office or agency does not constitute a waiver if the holder took reasonable steps to prevent disclosure and also promptly took reasonable steps to rectify the error.

FRE 502 ADVISORY COMMITTEE NOTES

- ❑ Depending 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.

About Tom Myrick

Member | Moore & Van Allen | Charlotte, NC

704.331.1126 | tommyrick@mvalaw.com

<http://www.mvalaw.com/professionals-608.html>

Tom Myrick has a wide range of civil trial experience, including frequent appearances in both state and federal trial and appellate courts, with particular experience in business related litigation.

He has been first chair trial counsel through verdict in dozens of civil jury trials, including trials that lasted up to four months. His practice has recently been focused on handling significant commercial cases involving sophisticated damage theories.

Mr. Myrick also counsels clients engaged in complex business combinations on methods to minimize their exposure to litigation. He speaks regularly to groups of trial attorneys around the country about ongoing developments in the law. He has spoken to audiences as far away as Japan where he addressed managing product liability issues on behalf of a firm client whose product is distributed internationall.

Practice Areas

- Business torts
- Contract disputes
- Corporate governance issues
- Franchise class action claims
- Product Safety Claims
- RICO claims
- Securities fraud
- Shareholder derivative claims
- Tax disputes
- Unfair business practices

Of Note

- Certified Public Accountant for more than 20 years (currently inactive)
- Trustee since 1997 for several charitable remainder trusts with multi-million dollar portfolios
- AV peer rated in Martindale-Hubbell legal directory
- Martindale-Hubbell Top Rated Lawyers: Commercial Litigation, 2012
- Included in Best Lawyers in America for Commercial Litigation, 2011-2014 and in Real Estate Litigation, 2012- 2014.
- Selected for inclusion to the North Carolina Super Lawyers list in 2007. His primary area of practice is Business Litigation.
- Recognized in the 2010-2014 editions of Benchmark Litigation as a "Local Litigation Star" in the area of Commercial Litigation
- Selected as a North Carolina Top Rated Lawyer by Martindale-Hubbell
- Member, Board of Directors, and Chair, Justice Initiatives, Inc.

Education

- B.S.B.A., University of North Carolina at Chapel Hill, 1978
- J.D., University of North Carolina at Chapel Hill, 1983