SQUIRE
PATTON BOGGS

New York State Bar Association
The Dispute Resolution Section
Benjamin N. Cardozo School of Law

Commercial Arbitration Training for Arbitrators and Counsel:
Comprehensive Training for the Conducting of Commercial Arbitrations
June 2, 2015
2:45 p.m.

Electronic Discovery: Procedural & Evidentiary Issues

Search & Review Issues/Predictive Coding

Richard L. Mattiaccio

Why Do We Care About Search & Review?

- As data populations increase, review costs increase

- Need to be able to review in a cost-effective manner to handle cases in a cost-effective manner

- Need to certify discovery responses pursuant to FRCP Rule 26(g)

SQUIRE◆
PATTON BOGGS
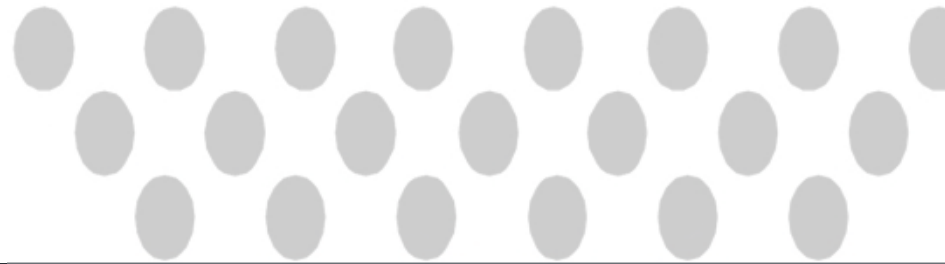
Date, de-dupe, de-NIST,
keywords with analysis,
source

Bulk Culling

Domains, author/recipient,
likely/relevance/privilege,
clustering

Intelligent Culling

Automation,
law firm or contract

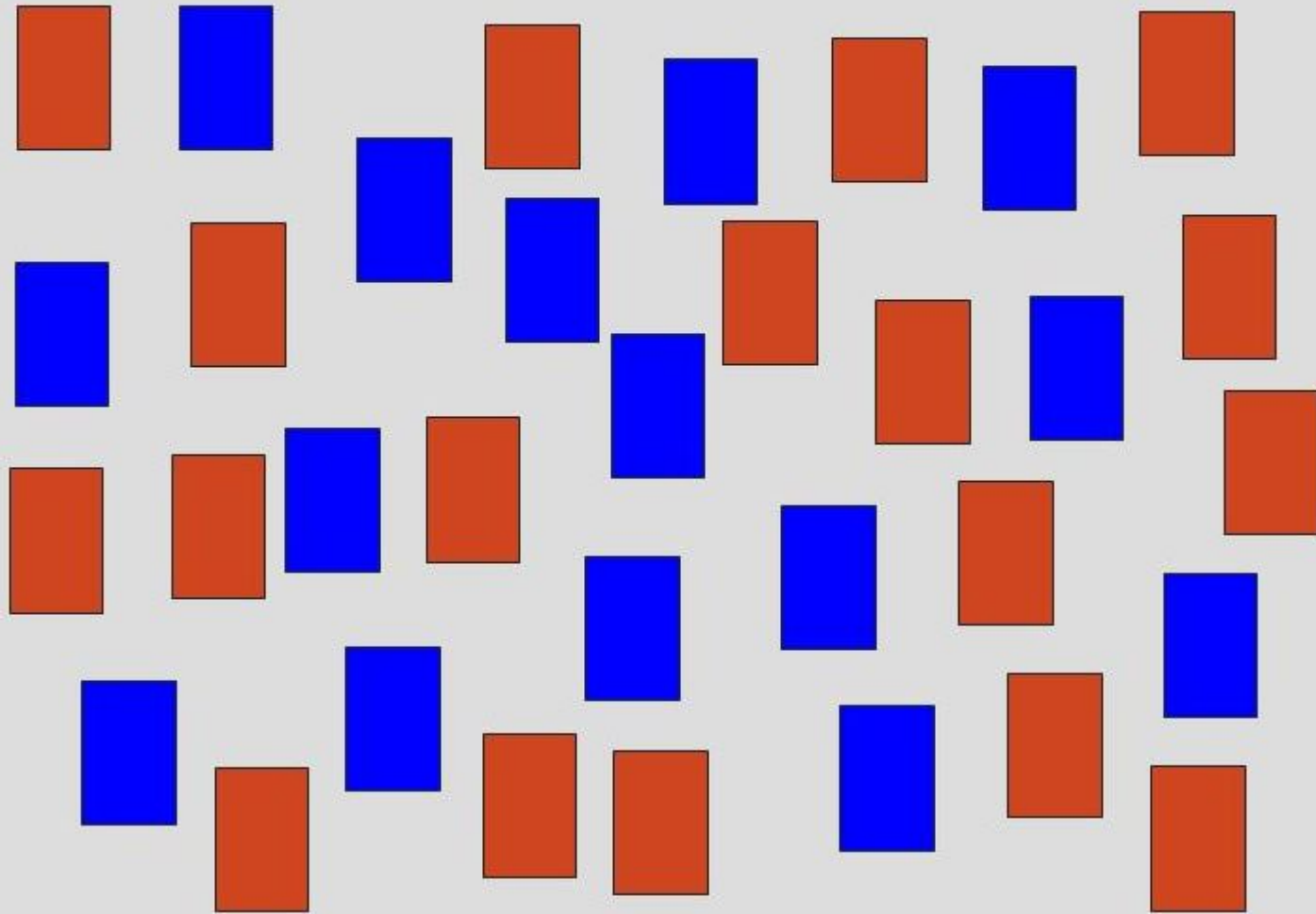**Relevance Review**

Automation,
reasonableness, clawback

Privilege Review

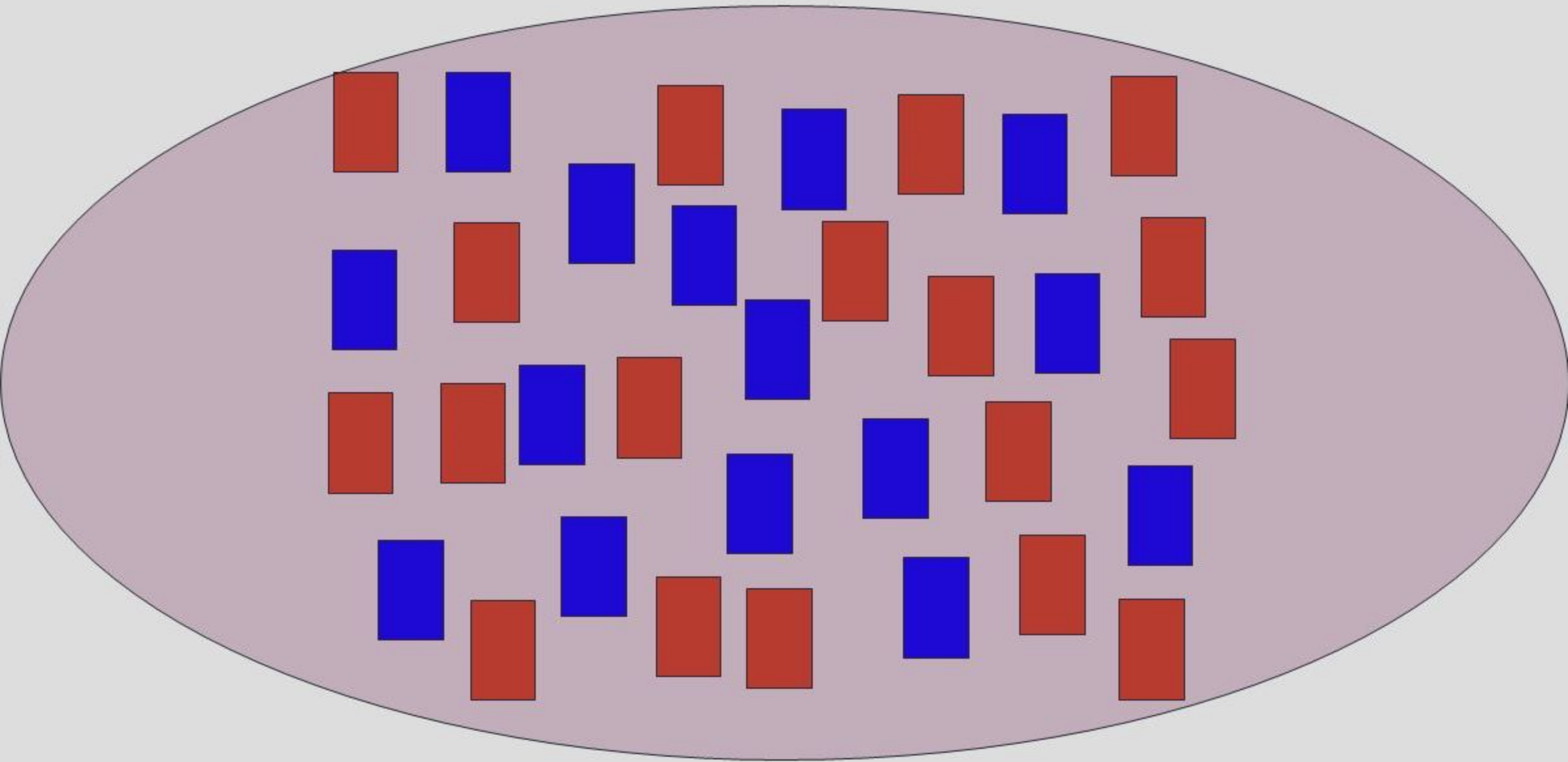Format, privilege log,
quality control
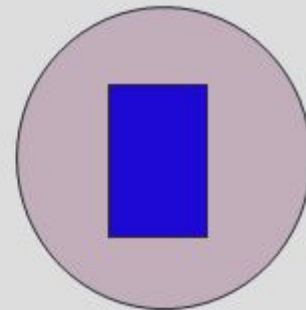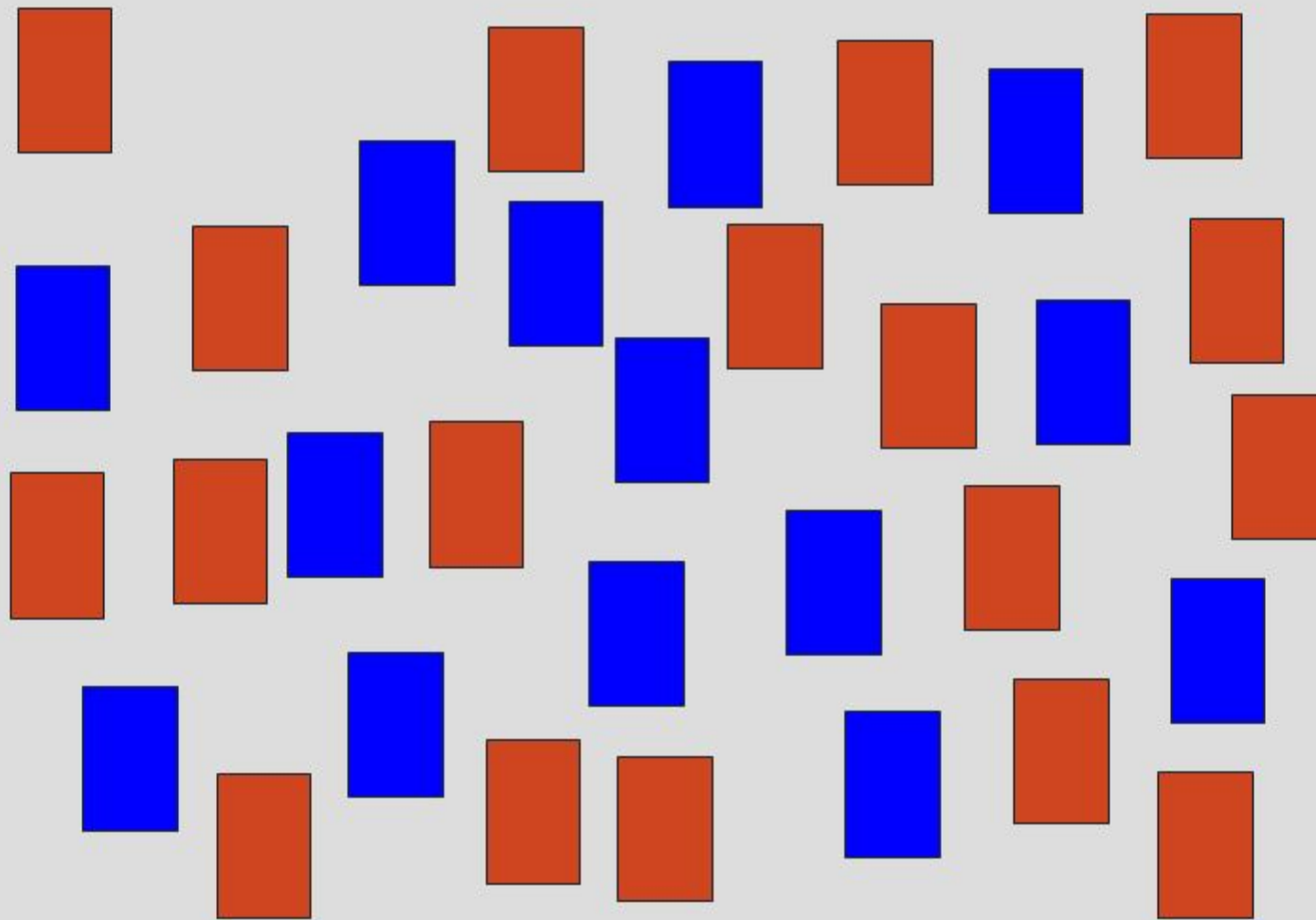
Production

Recall / Precision

Recall / Precision

Recall

Recall / Precision
Precision

Recall / Precision
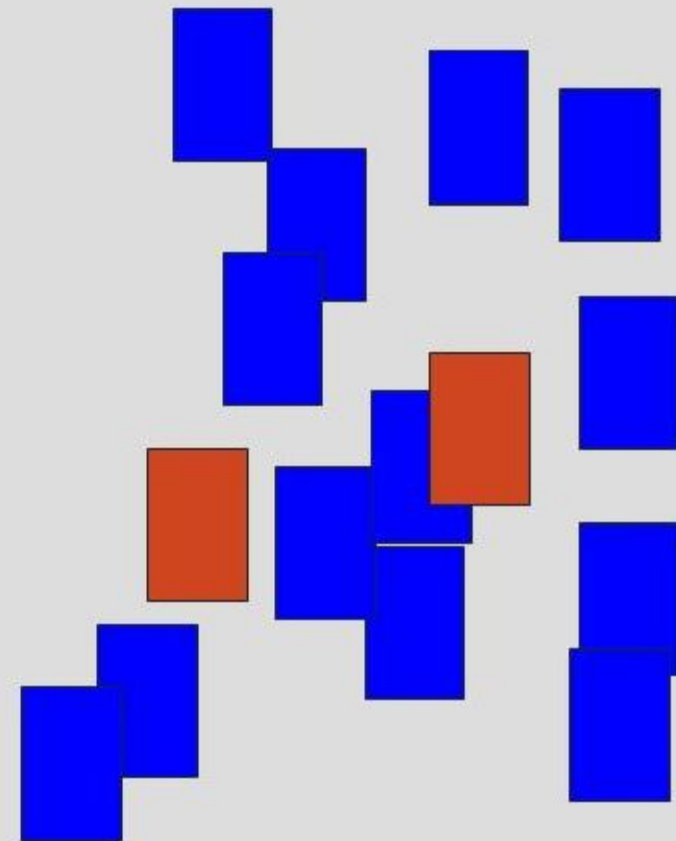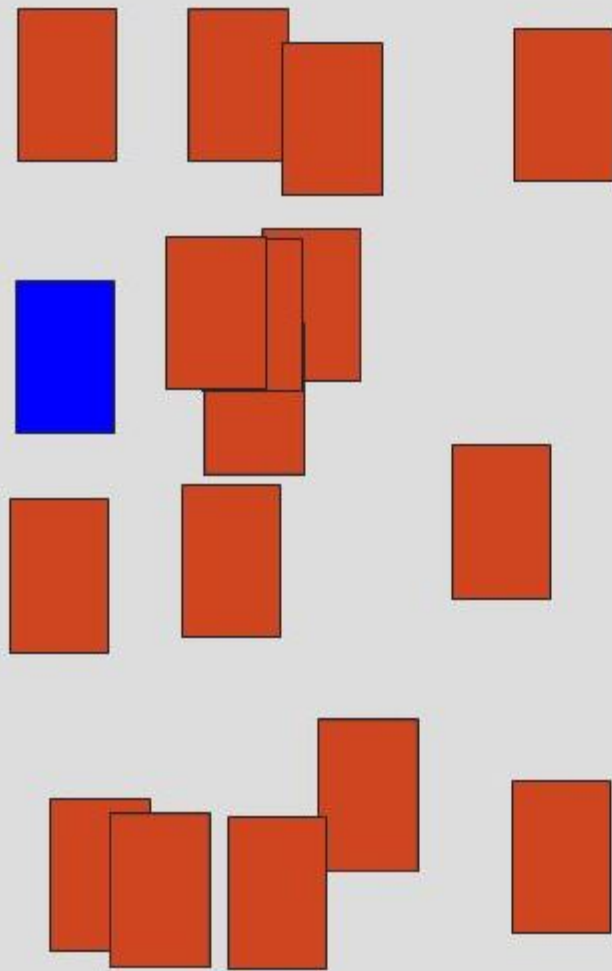
Issues with Keyword Searching

- Overinclusive and underinclusive

- May fail to account for misspellings, abbreviations, and acronyms

- "False Confidence Syndrome" – false confidence arising from the relatively high precision of keyword search results where it is not apparent that recall is low

- Common and easily understood search method

- Well-crafted keywords can be an acceptable method

Judicial Commentary on Keywords

- "There are significant issues with keyword searches. Keyword searches work best when the legal inquiry is focused on finding particular documents and when the use of language is relatively predictable."

  - *In re Direct Southwest, Inc.,* 2009 U.S. Dist. LEXIS 69142, at *5-6 (E.D. La. Aug. 7, 2009)

- "While keyword searches have long been recognized as appropriate and helpful for ESI search and retrieval, there are well-known limitations and risks associated with them, and proper selection and implementation obviously involves technical, if not scientific knowledge."

  - *Victor Stanley v. Creative Pipe*, 250 F.R.D. 251 (May 2008)

Qualifications to Craft Search Queries

- Whether search terms or "keywords" will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics. . . . Given this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread. This topic is clearly beyond the ken of a layman. . . .

  -*United States v O'Keefe,* 537 F. Supp. 2d 14, 24 (D.D.C. 2008)(Facciola, J.); *Equity Analytics, LLC v. Lundin,* 248 F.R.D. 331, 333 (D.D.C. 2008)(Facciola, J.)

"Intelligent Discovery"

- Use of technology and workflow techniques to cull relevant and privileged documents
- Potential to save hundreds of thousands of dollars of review costs
- Proven to be better than straight keyword searching
- Defensible process – bolstered by internal evaluation and testing

Predictive Coding Evaluation

- Beginning late 2009, Squire Sanders conducted a study of predictive coding software

- Actual data from previously reviewed collection was used for study

- Results demonstrate that this technology is at least as effective, if not more effective, than human review at identifying potentially relevant documents

Evaluation Results

- After an 11 hour training process, 44,000 documents were scored by the system.

- Based on a sample of ~3500 documents previously reviewed by a Squire Sanders team, system coding agreed with the attorney (human) review 87% of the time

Setup	Interactive Ranking	Sample Results	Batch Ranking	Final Results	Utilities

Discrepancy Matrix

		Predictive Coding		
		Relevant (X)	Non-Relevant (Y)	Total
Reviewer	Relevant (A)	1274	124	1398
	Non-Relevant (B)	329	1771	2100
	Total	1603	1895	3498

13

Discrepancy Analysis

- 100 of the non-agreement documents were sampled.
- Results indicate that the predictive coding system performed as well as the Human Review team.

Expert Verification

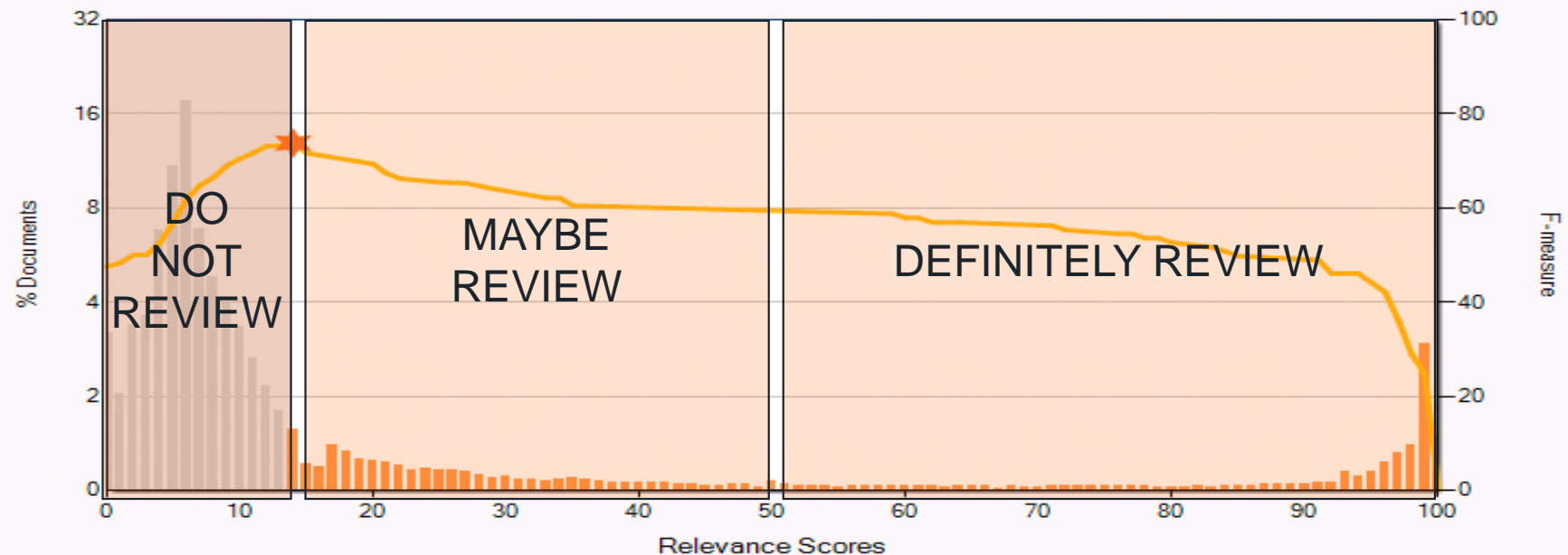|	Verification	
	Sample Size	**Actual Relevant**
Reviewer Relevant / Predictive Coding Non-Relevant	50	16
Reviewer Non-Relevant / Predictive Coding Non-Relevant	50	21

Results

	Recall	Precision	F-Measure	Accuracy
Predictive Coding	94.1%	88.1%	92.8%	92.0%
Reviewer	89.8%	96.4%	91.0%	94.2%

Leveraging Predictive Coding Scoring

Leveraging Predictive Coding Scoring

Relevance Distribution Graph

Contract Staffing

Jr. Associate/ Paralegal

Senior Attorney

% Documents

F-measure

Relevance Scores

Culling Statistics

Culling Percentage:	74	
Cutoff Score:	14	
Estimated Recall:	84.9%	(Of the relevant documents, how many did Predictive-Coding find?)
Estimated Precision:	47.36%	(Of the documents Predictive-Coding considered relevant how many are truly relevant?)
Estimated Fmeasure:	73.76%	(A measure of the balance between recall and precision)

☐ Filter By Keywords: — All —

16
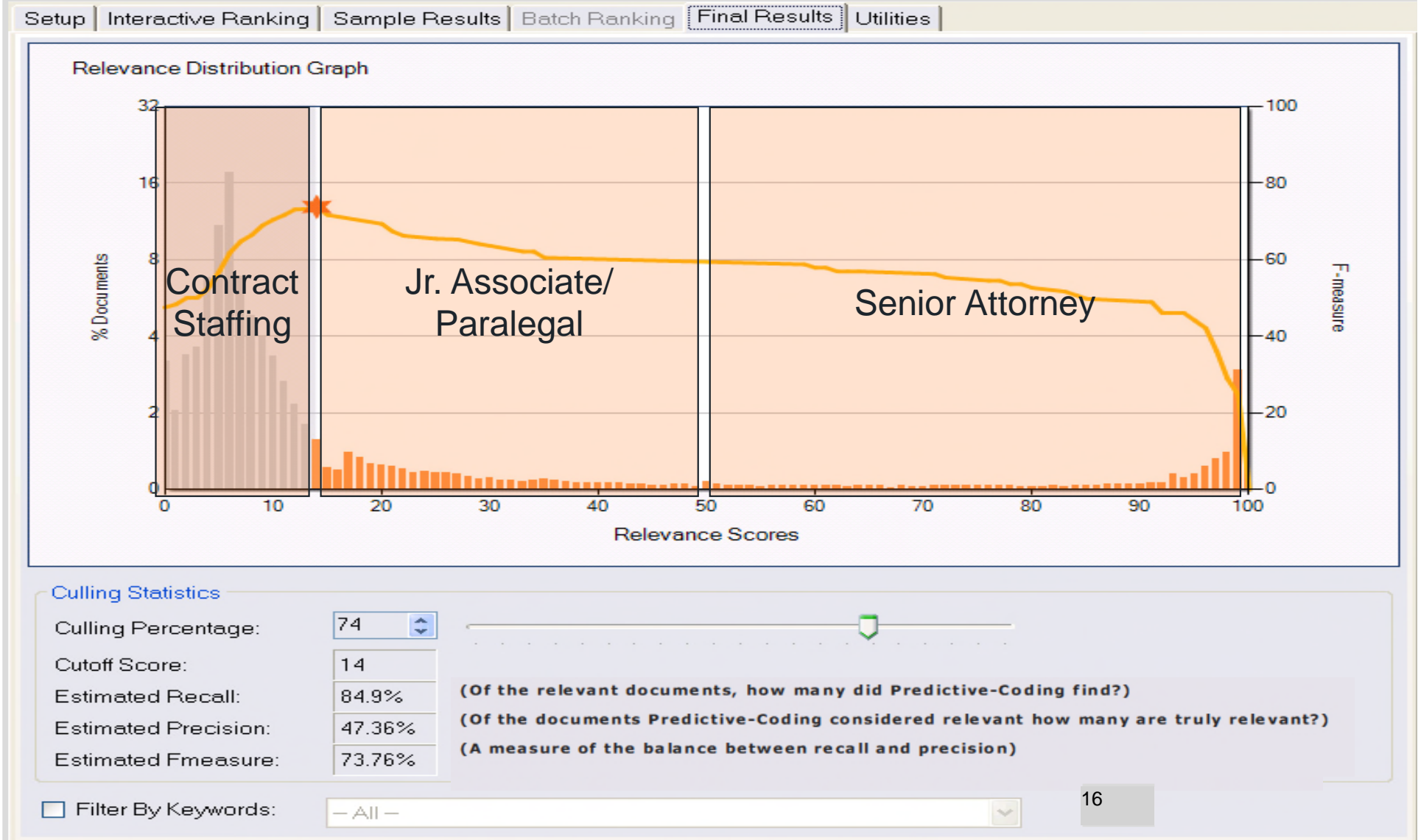
Review Budgeting

TOTAL DATA – 200 GIGABYTES   **Est. 1,400,000 Documents**

	Reduction%*	Removed	Remaining GB
De-NIST	15%	30.0	170.0
File Type Filter	10%	17.0	153.0
Custodian Filter	0%	0.0	153.0
Du-Duplicate	30%	45.9	107.1
Date Filter	15%	16.1	91.0
Keyword filter	30%	27.3	63.7
Domain analysis	10%	6.4	57.4
Bulk 'junk' removal	5%	3.2	54.2

* Estimate % data reduction and resulting impact

REVIEW OPTIONS

OPTION 1 – LINEAR REVIEW		
EST. DOCS TO REVIEW	LAW FIRM*	CONTRACT*
379,161	$889,816.38	$205,660.56
* INCLUDES ADDITIONAL COSTS OF ECA, REVIEW SUPERVISION, QC AND KNOWLEDGE MGMT		

OPTION 2 – PREDICTIVE CODING CULL		
LOAD COST	TRAINING/QC COST	ECA/KM
$37,916.08	$17,550.00	$27,000.00
TOTAL COST OF PREDICTIVE CODING CULL		$82,466.08

Current Issues with Predictive Coding

- Increased judicial acceptance

- Lingering confusion about how to properly apply the technology (seed sets, training, etc.)

- Debate about required level of transparency

- Will continue to gain traction and become routinized based on economics

Selected Reading and Resources

Cases

- *Da Silva Moore v. Publicis Groupe SA*, No. 11 Civ. 1279 (ALC) (AJP) (S.D.N.Y. Apr. 25, 2012)

- *Nat'l Day Laborer Org. Network v. United States Immigration & Customs Enforcement Agency*, --- F. Supp. 2d ---, 2012 WL 2878130 (S.D.N.Y. July 13, 2012)

- *Brookfield Asset Mgmt., Inc. v. AIG Fin. Prods. Corp.*, No. 09 Civ. 8285(PGG) (FM), 2013 WL 142503 (S.D.N.Y. Jan. 7, 2013)

Articles

- *Best Practices in E-Discovery in New York State and Federal Courts*, and, in particular, its Guideline No. 12 & Comment, Report of the E-Discovery Committee of the Commercial and Federal Litigation Section of the New York State Bar Association (July 2011) http://www.nysba.org/Sections/Commercial_Federal_Litigation/ComFed_Display_Tabs/Reports/ediscoveryFinalGuidelines_pdf.html

- *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient than Exhaustive Manual Review*, M. R. Grossman & C. V. Cormack, XVII Richmond Journal of Law and Technology 11 (2011), http://jolt.richmond.edu/v17i3/article11.pdf

- *Evaluation of Intelligent Discovery Processes,* H. Nicols & S. J. Goldstein, 27 Legal Tech Newsletter 9 (January 2011), http://www.lawjournalnewsletters.com/issues/ljn_legaltech/28_11/news/154735-1.html