Predictive Coding

Technology Assisted Review

Frederick N. Kopec
General Counsel
Scarab Consulting

SCARAB CONSULTING

Today's Speaker: Frederick N. Kopec, J.D.

- General Counsel for Scarab Acquisition, LLC, d/b/a Scarab Consulting
- Former Vice President and General Counsel for Renew Data Corp.
- Former Deputy Attorney General, Litigation Division, State of Indiana (pre-electronic evidence era)
- Over 30 years of legal experience, including 25 years of corporate practice
- Graduate of the University of Notre Dame and the Indiana University Maurer School of Law

SCARAB CONSULTING

Technology-Assisted Review
What is it?

Technolgy-Assisted Review ("TAR"): a broad term that means what it says: using a technology to assist review.

The technology typically refers to some form of language-based analytics.

SCARAB CONSULTING

Predictive Coding
What is it?

- Predictive coding is a form of technology-assisted review
- Uses classifiers to extrapolate human coding decisions made on a subset of materials to a larger data set, requiring an iterative approach that includes statistical sampling and QA to refine and improve the classifer
- Offers computer-generated document relevance rankings
- Those relevance rankings ARE the predictive coding
- Relevance rankings have been available in other software applications such as spam filters and Pandora's music service, for some time now, so it is not a new technology
- What is new here is the formal workflow incorporating analytics, such as relevance ranking
- It is NOT concept clustering or a keyword search

SCARAB CONSULTING

Court Acceptance-Computer-Assisted Review

"Computer-assisted review now can be considered judicially-approved for use in appropriate cases."

Judge Andrew Peck, United States Magistrate Judge, Southern District of New York, Monique Da Silva Moore, et al., v. Publicis Groupe, et al., No. 11 Civ. 1279 (S.D.N.Y. Feb. 24, 2012).

SCARAB CONSULTING

What is an “appropriate case” for PC?

- Any matter of at least 100,000 documents (that could require 10 – 20 attorneys at least two weeks to review)
- Cases involving 50 or more custodians
- Class actions
- Regulatory matters (e.g. FTC or DOJ) that require large-scale review in short period of time
- Commercial litigation with similar or consistent data types
- Multi-lingual cases (software classifier must have multi-lingual capabilities)

SCARAB CONSULTING

What is not an “appropriate case” for PC?

- Small, single-plaintiff cases
- Cases with few custodians and limited data sets
- Cases with large populations and small number of responsive documents (e.g. employment cases)
- Cases with a high number of documents that are NOT text-based (e.g. photographs, images, audio files, spreadsheets, drawings), such as patent litigation cases

SCARAB CONSULTING

Recent Cases – Predictive Coding

- Da Silva Moore v. Publicis Groupe, 2012 WL 607412 (S.D.N.Y., Feb. 24, 2012)
- Global Aerospace v. Landow Aviation, No. CL 61040 (Va. Cir. Ct., Loudoun County, April 23, 2012)
- Kleen Products v. Packaging Corp. of America, No. 10-C-5711 (N.D.IL)
- In Re: Actos (Pioglitazone) Products Liability Litigation, W.D.LA (MDL No. 6:11-md-2299) ESI Protocol dated July 27, 2012

SCARAB CONSULTING

Da Silva Moore v. Publicis Groupe

- US Magistrate Judge Andrew Peck approves the use of Predictive Coding to aid in the production and review of 3 million electronic documents.
- Judge Andrew Carter, Jr., adopted Peck's decision on April 25, 2012
- "Computer-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."

SCARAB CONSULTING

Global Aerospace v. Landow Aviation

- The local court approved a predictive coding protocol after defendants moved for a protective order due to the volume of data sought by plaintiffs
- Relies heavily on *Da Silva Moore*
- Plaintiff's Memorandum in Support of Motion for Protective Order Approving the Use of Predictive Coding is a textbook for the practitioner who wants to convince a court of the value and statistical merit of Predictive Coding

SCARAB CONSULTING

Kleen Products v. Packaging Corp. of America

- A federal antitrust case
- Plaintiff sought to force defendant to use Predictive Coding to ensure the accuracy of defendant's document production
- The arguments in the evidentiary hearings pitted eDiscovery experts against each other in a Predictive Coding vs. Keyword Search battle
- Judge Nolan's frustration with the parties is evident in the hearing transcripts – "Why can't we all just get along?"
- No clear win for Predictive Coding

SCARAB CONSULTING

In Re: Actos (Pioglitazone) Products Liability Litigation

- Federal litigation in Lafayette, Louisiana, involving drug manufacturer Takeda Pharmaceutical's oral anti-diabetic medication "Actos"
- The parties' Case Management Order: Protocol Relating to the Production of Electronically Stored Information ("ESI") is an extremely detailed example of an agreed-upon Predictive Coding protocol
- The details of the protocol are driven by the software used by the parties (Equivio's Relevance)

SCARAB CONSULTING

Predictive Coding Protocols – *Da Silva* Workflow

- Software = Recommind Axcelerate
- 1. Identify the initial seed set (using "search and analytical tools, including keyword, Boolean and concept search, concept grouping, and up to 40 other automatically populated filters available within the Axcelerate system").
- 2. Seed sets tested with keywords submitted by both sides to create a "high priority relevant seed set".
- 3. Apply seed sets to the Predictive Coding process, "training" the software.
- 4. Trained software then identifies and prioritizes all substantively similar documents in the data collection.
- 5. Attorneys do QA check by hand reviewing and coding a judgmental sample of at least 500 documents to ensure their proper categorization and to further calibrate the system by recoding documents into their proper categories. Repeat iterative process.
- 6. At the end of the iterative process, all documents predicted by Axcelerate to be relevant will be manually reviewed for production.
- 7. The accuracy of the search processes, both the systems' functions and the attorney judgments to train the computer, will be tested and quality controlled by both judgmental and statistical sampling.

SCARAB CONSULTING

Predictive Coding Protocols – *Global Aerospace* Workflow

- Software = Unknown

- 1. Defendant Landau to train the Predictive Coding tool, then provide all documents used to train the software (except for privileged and sensitive irrelevant documents) to Plaintiff's counsel.
- 2. Predictive Coding tool used to categorize all documents being reviewed. A "statistically valid sampling program" will be used "to establish that a majority of the relevant documents have been retrieved." Defendant proposes retrieval of 75% of the relevant documents as the acceptable retrieval rate.
- 3. Achievement of at least a 75% level of recall (percentage of relevant documents retrieved) equals an acceptable result.

SCARAB CONSULTING

Predictive Coding Protocols – *In Re: Actos* Workflow

- Software = Equivio Relevance
- Consultant = Epiq Systems
- 1. Epiq to conduct sample collection from four of defendant's key employee custodians. Other regulatory documents added to this sample collection population.
- 2. Each side to nominate three "experts" to work collaboratively to train the Equivio Relevance system, who will first be trained on the Equivio Relevance software and coding process.
- 3. The Experts work together to make a single "relevance" decision for each document in a subset of the sample collection population. The Equivio Relevance system is trained until the results it produces reach Stability.
- 4. The System then calculates "relevance" scores for each document in the entire sample collection population, giving each document a relevance score of 0 – 100.
- 5. No seeding will take place.

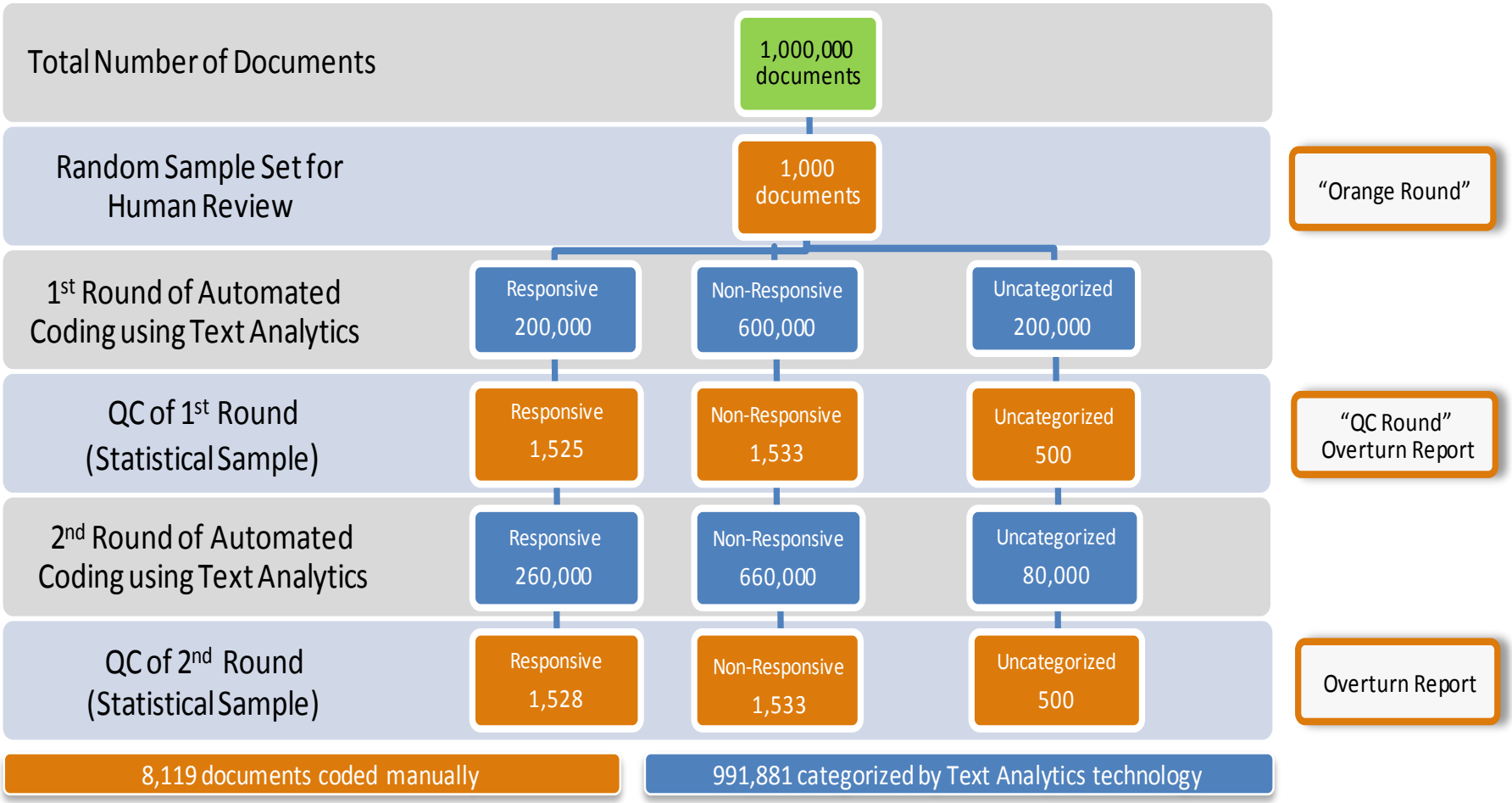
SCARAB CONSULTING

Predictive Coding Protocols – *In Re: Actos* Workflow, continued

- 6. An "Active Learning" process will take place in which the System chooses sample documents for the experts to review, and chooses the next set of samples based upon how the experts code the previous sample set. The System determines (through a non-configurable setting) that it has reached Stability (and is ready to begin Predictive Coding) when the marginal contribution of additional samples to the enhancement of the System's ability to classify the documents approaches zero.
- 7. Parties agree upon the minimum relevance score, above which all documents will be manually reviewed for production. Predictive coding occurs, and all non-privileged relevant documents resulting from the manual review will be produced for Plaintiffs.
- 8. A random set of documents below the minimum relevance score is reviewed by the experts to verify that these non-produced documents contain a low prevalence of relevant documents.

SCARAB CONSULTING

Scarab Consulting Proficient Review™ Summary Workflow



Notable (Non-Blog) Articles

"Overall, the myth that exhaustive manual review is the most effective – and therefore, the most defensible – approach to document review is strongly refuted. Technology-assisted review can (and does) yield more accurate results than exhaustive manual review, with much lower effort."

Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review,* Richmond J. of Law & Tech., Vol. 17, No. 3, (2011).

"The use of the kind of processes employed by the two systems in the present study can help attorneys to meet the requirements of Rule 1 of the Federal Rules of Civil Procedure: ' to secure the just, speedy, and inexpensive determination of every action and proceeding.'"

Herb Roitblat, Anne Kershaw, and Patrick Oot, *Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review,* Journal of The American Society for Information Science & Technology, Vol. 61, No. 1 (2010).

SCARAB CONSULTING

How to Get Court Approval for Predictive Coding

The other side won't agree. How do you persuade your judge to implement a predictive coding protocol in your case?

- **Know your data** inside and out.
- **Keep it simple.** Compare your alternatives, and explain the consequences of paper review or keyword searches.
- **Be open.** Seek agreement on a protocol. If you must seek court intervention, propose an open, transparent process.
- **Rely on common sense legal principles.** Spending more on discovery than the amount in controversy is always a bad idea.
- **Never promise perfection.** Predictive coding will not solve every EDD problem.

Thanks to Peter Buckley and Scott Vernick
Law Technology News, August 15, 2012

SCARAB CONSULTING

Thank you very much!

Frederick N. Kopec

General Counsel

Scarab Acquisition, LLC

fkopec@consultscarab.com

512-448-3005

SCARAB CONSULTING