ITERATIVE LEGAL ANALYSIS & SAMPLING ("ILAS") vs. LINEAR DOCUMENT REVIEW

– A COMPARATIVE CASE STUDY

Can lawyers conducting iterative data analysis and sampling identify relevant documents more efficiently than legal teams undertaking linear review?

By Anne Kershaw, Esq.

Adjunct Professor, Masters of Science, Information and Knowledge Strategy

Columbia University School of Professional Studies

sps.Columbia.edu

COLUMBIA UNIVERSITY
School of Professional Studies

June 2016

Acknowledgements

A special thanks to the many talented people who donated time and resources to this important study, including:

- Nathanial Byington of Zoellner Byington, who provided the statistical analysis;

- Kelly Services, who provided the document reviewers;

- Matthew Golab, who performed the analytics review;

- Brian Seward, who managed the study;

- Jonathan Swerdloff, who is drafting a law review article about the study;

- Reasonable Discovery, LLC, who provided the data and the servers; and

- Jo Sherman of EDT, who donated the software for both the Iterative Legal Analysis and Sampling Team and the linear review team.

Table of Contents

The Hypothesis

Can an experienced lawyer who understands the case, the client objectives and strategy, and possessing technical analysis skills or working collaboratively with someone who has them (an "Iterative Legal Analysis and Sampling" (ILAS) approach), find the relevant documents in litigation or an investigation, faster, less expensively and as effectively as, a team of contract review lawyers working at much lower hourly rates conducting 'linear'[1] document review?

The Answer

Yes. The study described below (the Study) determined that, using the same software, an ILAS approach can find all of the relevant documents found by a linear review and also many documents missed by a linear review, in a fraction of the time and for a fraction of the cost. In other words, a subject matter expert armed with iterative analysis and sampling tools can perform better than a team of linear reviewers.

The Need for This Study

Traditionally, all the documents relevant to a dispute or investigation were contained in paper based files, filing cabinets or in boxes in a warehouse. Lawyers were required to read all these documents because there was no other way to understand the facts of the case, the information at hand and what might need to be produced to satisfy discovery obligations.

With the emergence of photocopiers, fax machines and word processors, more and more paper was produced. So, more and more paper files were collected, as were more and more boxes in the warehouse. Hundreds of boxes instead of a handful. The only way to find what was responsive was to arrange for trained human eyes to read it all.

---

[1] Linear review is the process of having human eyes on all documents within the review universe.

Clients complained about the cost of experienced litigators eyeballing every document and this was the impetus for the emergence of a new document review industry. The answer to the document volume and cost problem was to push the review process to less expensive lawyers interested in temporary project work. At the time, this was really the only viable option. There was simply no other way to understand the documents, explore the facts, and hone in on the potential evidence.

Then came electronic documents and email and e-discovery volumes increased dramatically. The contract linear review lawyers manually reviewed these electronic documents as well, just as they had done for paper documents, but the volumes were exponentially larger, leading to enormous costs for clients and enormous profits for e-discovery vendors and law firms.

Many of us working on reducing the costs of e-discovery have long questioned the continued use of linear review in light of increasingly powerful electronic search capabilities for iterative sampling and analysis. The theory that we can find what we need to find if we understand what we have and how to use the technical tools at our disposal, is now finally, definitively, proven to be true. In light of these findings, linear document review can no longer be considered best practices for most cases. In some instances, a blended process applying ILAS before engaging a linear review may be most appropriate and would also reduce document review costs.

The Framework for the Study

The Study compared two methodologies to locate relevant documents. A traditional linear document review was performed by two document reviewers, with oversight by a quality control reviewer. The results of the linear review were then compared with an "Iterative Legal Analysis and Sampling" (ILAS) approach.

The ILAS approach involved a senior legal technology manager and a senior attorney familiar with the matter, using software with integrated review and analysis functionality, iteratively applying a range of strategic and creative search techniques to

find responsive documents. The ILAS approach did not involve a 'document review' in the traditional sense. Instead it involves a more skillful application of human expertise using metadata filters (date ranges, file types, authors and recipients etc.), correspondence traffic, domain and name analysis tools, alongside keywords that were refined iteratively by viewing the search terms in context within the documents that contained them during a circular, feedback-loop refinement process.

The discrepancies between the two review processes were then reviewed by the counter process and "agree" and "disagree" determinations were counted and resolved.

The Documents

The documents used for the study came from a real case filed against a software systems company by a former client that was settled shortly after data was collected. Collections had been made from 12 custodians and totaled 9.95 gigabytes containing 113,454 documents after the removal of system files. The data set was de-duplicated across custodians, with all source information retained, and then keyword culled to a final review set of 9764 documents.

The document types in the final review set included email, Microsoft Office documents, PDFs, and other assorted documents. Certain document types were unique to the systems in place at the client, such as reports from databases.

The Teams

The linear review team consisted of two contract attorneys contributed by Kelly Services, a traditional document review staffing agency. The Kelly Services team was hand selected for their high quality work on prior matters and for their extensive document review experience. Their work was overseen and quality checked by an experienced, senior document review attorney who also managed the project. The ILAS team included a senior Legal Technology Manager who is an expert in the use and

application of document review analytics and a senior attorney from [Reasonable Discovery, LLC](#).

The Review Scope

The scope of the review was framed by the claims made in the underlying case and the document requests served in the case. Both teams were provided with copies of the pleadings and document requests. The review teams were instructed to mark as relevant any documents that pertained to modifications, alterations, improvements or repairs to any of the defendant software developer's goods or services. Documents were also deemed relevant if they were related to issues or problems with, or defects existing in, any of the defendant's goods or services.

The Review Platform

Both the linear and ILAS teams coded all 9764 documents using [EDT](#), a document processing and hosting repository that provides both a traditional document review interface and analysis functionality to facilitate the ILAS approach.

The linear team used the traditional document interface to make document by document decisions as to relevance. The ILAS team was given access to the document analysis functionality within the platform. Each team coded the documents independently of the other team and EDT's security settings were used to suppress from view the coding decisions made by the other team.

The Linear Review

The first pass team divided the documents randomly into three folders – one for each first level reviewer so that each reviewer had a similarly-sized cross section of all the documents in the set. The first pass reviewers started on their documents after a training on the subject matter and reading through the pleadings and document requests. Document families were used to give context to document meaning but each document was coded individually. The only coding in this instance was a binary

relevance coding for responsive or not responsive to the issues outlined above. No coding was applied for privilege, confidentiality, or issue tags.[2]

As the review started, the reviewers were encouraged to talk amongst themselves regarding the meaning of documents. After they reviewed their first 200 documents, they rotated computers and checked through another reviewer's documents to ensure both familiarity with the documents and conformity of the decisions. Any feedback was openly shared and further discussion was encouraged among the group. After concluding the check of each other's documents, the reviewers switched back to their original computers and continued to review their original folder of documents.

Throughout the review, the reviewers elevated questions on the interpretation of documents to the senior attorney project manager, who would make a final decision on the coding of the documents. This decision was communicated to all of the reviewers to apply going forward. A note was made in a decision log regarding the responsiveness determination for purposes of quality control across the entire review set. A quality assessment of the review was performed after the linear review was complete. Leveraging the decisions recorded in the decision log and the experience of the senior attorney, targeted searches were used to correct any coding discrepancies.

The first pass review concluded in 4 days of review. The quality assessment review finished in a few days following the first pass review.

The Iterative Legal Analysis and Sampling (ILAS) Approach

The ILAS team leveraged EDT's integrated document analysis functionality and review features to allow for iterative metadata analysis and in-depth, real time searching, together with random document content sampling. Specifically, the analytics

---

[2] Another study is in the works to measure the effectiveness of ILAS for identifying privileged and confidential documents versus traditional methods.

reviewer focused on keyword searches using stemming and Boolean searches. These terms were continually refined via real time analysis of documents containing keyword hits. Email to, from, and between specific custodians were searched and email addresses associated with the individuals involved on each side of the underlying suit, and emails between them, were identified and deemed relevant. Emails threads were clustered and similar documents were programmatically identified and presented. Other metadata searches included searching for specific email subjects using email address and domain normalization techniques. In other words, the ILAS approach involved a true and robust technology assisted review, emphasizing human expertise and employing a strategic process that is very different from the process referred to as "predictive coding."

Document sets returned by the searches and filters were sampled at a rate of 10 - 20% in order to ensure that the analysis returned expected results. In instances where false hits were returned, keyword searches were revised and expanded iteratively to further refine results. This winnowing process leveraged the functionality of the application's multiple analysis tools, allowing the ILAS team to quickly find the correct documents for further analysis. The analytics reviewer was able to confirm his analysis with a lawyer who had extensive experience with the documents through prior work on the underlying case.

The ILAS team took a total of 10 person-hours to complete and that time period included certain technology associated times as documents were re-indexed. In addition to the 10 person-hours required for the review, there were 4 hours of attorney supervision performed as oversight.

Coding Disparities and Adjudication

The linear review coded 4894 documents as responsive and 4870 as non-responsive. The ILAS reviewer coded 6304 documents responsive, 3460 non-responsive.

Initial Pass

	Linear Team	ILAS Team
Documents coded Responsive	4894	6304
Documents coded Non Responsive	4870	3460
Time Spent	98 hours	14 hours

In addition, 1712 documents were coded responsive by the ILAS review team that were coded non-responsive by the linear review team, while 303 documents were coded responsive by the linear team that were coded non-responsive by ILAS team. These disagreements in coding were adjudicated by having each team review the other team's different coding decision. Here are the results:[3]

	Gold Standard	Linear Team	ILAS Team
Relevant	5448	4847	6225
Not Relevant	4190	4791	3413
Excluded as Borderline	126	126	126
Total Documents	**9764**	**9764**	**9764**
True Positives (correctly assessed as relevant)		4824	5215
True Negatives (correctly assessed as not relevant)		4167	3181
False Positives (incorrectly assessed as relevant)		23	1010
False Negatives (Incorrectly assessed as not relevant)		624	232
Excluded as Borderline		126	126
Number Relevant Documents Missed		624	232
Percentage Relevant Documents Missed		11.45%	4.26%
Relevant Documents found by ILAS team & missed by Linear Team			392
After ILAS Team's re-review of Linear Team decisions - final numbers 2			
	Gold Standard	Linear Team	Analytics Team
Relevant	5475	4847	6217
Not Relevant	4155	4783	3413
Excluded as Borderline	134	134	134

---

[3] The final coding decisions and results were analyzed by Nathaniel Byington, an expert in computational linguistics, for purposes of comparing the effectiveness of each review.

Total Documents		9764	9764	9764
True Positives (correctly assessed as relevant)			4824	5242
True Negatives (correctly assessed as not relevant)			4132	3181
False Positives (incorrectly assessed as relevant)			23	974
False Negatives (Incorrectly assessed as not relevant)			651	233
Number Relevant Documents Missed			651	233
Percentage Relevant Documents Missed			11.95%	4.28%
Relevant Documents found by ILAS Team & Missed by Linear Team				418

Conclusions

The ILAS team was able to find more relevant documents than the linear review team (a higher "Recall"), was less likely to miss a responsive document, and was more likely to include a borderline document, i.e. the ILAS approach was overly conservative in its assessment of responsiveness in line with the preference of most litigators. Where in doubt, include it. That approach is generally preferred in litigation to the alternative that leads to a high number of false negatives – relevant documents missed. The linear review took 98 hours whereas the ILAS review took 14 hours.

With these numbers, paying the linear review team members $50 per hour, the cost of the review was $4,900.00.[4] Paying the ILAS team $200 per hour, the cost would be significantly less at $2,800.00.

Observations and Closing Thoughts

The days of ineffective, expensive and time-consuming document review are finally over for most cases.[5] Now, finally having removed the barriers to ILAS, we can strategically, inexpensively and quickly find the relevant documents. Instead of doing

---

[4] This estimate does not include the cost of the more expensive senior quality control attorney or the time needed to train the document review lawyers.

[5] Some cases may benefit from a blend of ILAS and linear document review.

document review, contract lawyers can now be used to create timelines, witness kits, and deposition outlines, further reducing litigation costs. And some of these contract lawyers, if given the opportunity to learn and use the technology, will become ILAS experts, providing increased value while reducing document review costs. The added benefit for these legally qualified professionals is the potential to apply their expertise in more stimulating ways and to deliver more rewarding outcomes.

We can talk about precision and recall, statistics, process and workflows indefinitely, but in the end, a perfect review that is prohibitively expensive is not a good review and an inexpensive review that gets the wrong results is worthless. In this study, the ILAS team delivered a better result in that it found more relevant documents in a fraction of the time and for a fraction of the cost.

The growth of e-discovery no longer means the death of litigation on the merits – to the contrary, the strategic use of talented individuals and targeted technologies will allow clients to get to the merits faster and more effectively.

=================================

Anne Kershaw is an attorney and expert in e-discovery and legacy data management. In 2005 she conducted the first study comparing linear document review with technology assisted review, [Automated Document Review Proves its Reliability](). Kershaw published two subsequent studies in 2010, [Document Categorization in Legal E-Discovery: Computer Classification vs. Manual Review]() and [Mandating Reasonableness In A Reasonable Inquiry](). Kershaw is a managing director at [Reasonable Discovery, LLC](), is on the faculty of Columbia University's [Information and Knowledge Strategy]() Master's Program, is on the Advisory Board for the Cardozo Law School Data Law Initiative and the Georgetown Advanced E-Discovery Institute, and was a presenter at workshops on e-discovery at the Federal Judicial Center's 2010, 2011 and 2014 Conferences for Magistrate Judges.

About [EDT](): For over a decade, EDT Software has established an international reputation in relation to its full lifecycle software that manages e-discovery, investigations, and litigation. EDT has been embraced by a world-wide client base including GLOBAL 100 law firms, top tier accounting firms, corporations, government agencies, forensic experts, consultants and service providers. For more information, visit [http://www.discoveredt.com]().

About Reasonable Discovery, LLC: Reasonable Discovery, LLC consultants are strategists who use litigation expertise and mastery of current technologies to provide clients with flexible, powerful solutions for collecting, interrogating and managing data. Using data to its full advantage doesn't have to be complex or expensive. With Reasonable Discovery, every case and every client has a reasonable and formidable alternative. www.ReasonableDiscovery.com.