

DESI V: Workshop on Standards for Using Predictive Coding, Machine Learning and Other Advanced Search and Review Methods in E-Discovery

A Post-Conference Workshop at the
14th International Conference on Artificial Intelligence and Law
Casa dell' Aviatore, Rome Italy

14 June 2013

Session 3 Discussant Notes (Gil Keteltas)

- Dan Regard, [*A Re-Examination of Blair & Maron*](#)
- Eric Schwarz, [*The Structure of Predictive Coding: A Guide for the Perplexed*](#)
- Amanda Jones, [*Variability in Technology Assisted Review and Implications for Standards*](#)
- Hans Henseler, [*Semantic Search in E-Discovery: An Interdisciplinary Approach*](#)
- Dan Brassil, [*Toward a Meaningful E-Discovery Standard*](#)

Three themes emerge from this collection of papers: (1) task matters in e-discovery; (2) there appears to be a disconnect between what matters to lawyers and information retrieval scientists in search; and (3) variability in task and “what matters” should be accounted for in standard setting.

Task Matters

- **Dan Regard:** we should not blindly cite *Blair and Maron* for the proposition that keyword search is ineffective. The attorneys in the original *Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System* were satisfied with the results of their searches. But information retrieval experts – using measures of precision and recall – concluded they should not have been happy. Given the difference between information that is merely relevant under the Federal Rules of Civil Procedure and information that is meaningful in resolving disputed issues, “attorneys may stop and be satisfied with less than 100% or even less than 75% of the relevant documents” (p. 17).
- **Eric Schwarz:** there is a “bewildering” range of available predictive coding systems, but what we want *to do* with those systems matters in assessing their individual advantages and disadvantages. Is the user of the technology trying to accelerate review, trying to classify a population of data based on human labeling of a sample, or trying to do a little of both? Within predictive classification we have to focus on the *kind of sampling* (random, stratified or judgmental) that is optimal for the task at hand: “Our own view is that it’s worthwhile to assess, at each stage of predictive classification, what kind of sampling most effectively advances overall goals” (p. 11).

- **Amanda Jones:** “there is no single best recipe for optimizing TAR performance and no guarantee of equally strong results for every project. Instead, experimentation and customization are the best avenues for achieving optimal results.” There are numerous ways to deploy TAR and deployment should be tied to the goals and applications for the results sought (which can be many and varied).
- **Hans Henseler:** “data used in e-discovery is typically on a case-by-case basis, it can be noisy and is diverse in nature and origin.” Moreover, in any given case the discovery process is evolutionary: “Most of the time it is not exactly clear beforehand what is sought in the e-discovery setting, therefore search often starts exploratory.” Research should focus on the impact of semantic search approaches in a variety of “real” search settings. We can better understand precision and recall by focusing on specific problems and by allowing “end-users to intuitively and flexibly interact with documents and available information in increasingly large data sets.”
- **Dan Brassil:** standards should provide a means to measure precision and recall but not dictate minimum levels because it is the role of “practitioners, taking into account the specific goals and circumstances of the retrieval effort, and taking into account non-statistical data as well, to decide what levels are required to have confidence in the results of a review or retrieval effort.”

Disconnect Between Lawyers and Information Retrieval Scientists

- **Dan Regard:** *Blair and Maron* suggest that information retrieval scientists viewed the use of keyword search as a failure, but the lawyers who performed keyword searches were satisfied with the results. In my view this exposes a divide between that is critical to any standard setting exercise. One school of thought says the lawyers are wrong because they don’t have any idea what they missed, another says the scientists are wrong because they don’t have any idea what really matters.
- **Amanda Jones and Dan Brassil:** both advocate the ability to measure performance (although oppose the adoption of specific recall and precision numbers as standards). But this presents a struggle for lawyers who want to be able to demonstrate defensibility. While judges and parties are comfortable saying that perfection is not expected in e-discovery, there is a lot of handwringing by judges and parties once inevitable imperfection is realized.
- **Hans Henseler:** proposes to bridge, or at least better understand, these differences through an “interdisciplinary approach, spanning the fields of criminology, law, Information Retrieval and Natural Language Processing.”
- **Eric Schwarz:** also suggests a bridge between lawyer and IR specialist; he notes that predictive coding technology enables more meaningful discussions of

proportionality by “de-linking . . . volume and cost while maintaining or improving quality.” He suggests that predictive coding hypotheses can be tested and refined, even where data and the algorithms deployed vary.

Standard Setting – Personal Observations

These papers expose the challenge ahead if standards are to be adopted for e-discovery. While technologists are eager to have performance standards set for technology, in light of the variety of tasks lawyers face and the seeming disconnect between lawyers and scientists about how to define success, the form a performance standard might take is difficult to imagine.

In addition, given that we are still in the early days of predictive coding, machine learning, and other advanced search and review methods – at least in terms of deployment by lawyers in real world settings – performance-based standards might have the unfortunate effect of limiting innovation and encouraging “development to the standard.” As Amanda Jones and Jianlin Cheng note, “there is value in maintaining flexibility in TAR processes and . . . the industry should support ongoing innovation and creativity with regard to TAR implementation.”

At the same time, lawyers are hungry for process standards. A defensible *process* for the use of a range of advanced search and retrieval methods would likely do more to expand their use and acceptance by judges, lawyers and clients than performance standards. Of course, performance and process standards are not mutually exclusive and these papers show the need for lawyers and technologists to work and think together.