

## REFLECTIONS OF THE TOPIC AUTHORITIES ABOUT THE 2009 TREC LEGAL TRACK INTERACTIVE TASK

The seven Topic Authorities (“TAs”) that participated in the 2009 Interactive Task offer the following thoughts and observations based on their experiences with the TREC Legal Track:

1. **The Importance of Interaction Between Teams and TAs.** The TA role is intended to simulate the functions typically performed by a senior attorney overseeing a large document production. As were the TAs in the 2008 exercise, many of the 2009 TAs were surprised by the limited amount of contact and overall interaction that most teams had with their TAs. Very few of the teams used a significant portion of the ten hours they were each allotted to consult with their TAs. This led some TAs to suggest that the TREC Legal Track consider imposing a “minimum contacts” requirement. Moreover, the TAs’ surprise extended not only to the quantity of the communications, but also to their quality. Some teams did not appear to grasp how to formulate questions that would have helped them to clarify the topic’s scope or to arrive at a common conception of relevance. It was the general consensus of the TAs that the teams that understood how to generate such questions performed considerably better than those that did not, even if the team spent less time interacting with their TA, suggesting that ultimately, the *quality* of the team/TA interaction was more important than the *quantity*.
  
2. **The Importance of Early Study of the Document Collection.** Consistent with the observation of the 2008 TAs, the 2009 TAs suggested the need for teams to become familiar with the document collection early in the process in order to gain sufficient familiarity with the corpus to be able to formulate meaningful questions for discussion with the TAs. Teams that made such an early investment were better able to present their TAs with increasingly nuanced questions that allowed them to reach quicker and more precise agreement on the contours of relevance. Some of the 2009 TAs suggested that the TAs also be given direct access to the document set – or at least a random sample drawn from it – following their review of the complaint and requests for production, but before beginning their work with their teams. The TAs stressed the importance for *both* the TAs and the teams to develop an adequate baseline of knowledge about the topic, through research of publicly available sources, *as well as* through familiarity with at least some of the recurring themes in the document collection, in advance of their initial discussions. This observation would seem to comport with growing anecdotal support in the industry for what is commonly referred to as “early case assessment.” The TAs noted that, in addition to a general lack of familiarity with the document collection, few of the teams appeared to have engaged in any research beyond reading the complaint and requests for production they had been provided. Furthermore, many of the teams displayed little or no interest in participating in a “kick-off” call about the topic designed to replicate the training session typically provided to reviewers at the beginning of a large-scale document review. Some TAs suggested that the TREC Legal Track should consider requiring that all teams participate in such a “kick-off” call to establish a baseline definition of relevance. Given the geographical dispersion of teams and scheduling challenges, it was suggested that the “kick-off” calls might be taped so that

team members that were unable to participate live could listen to them at their convenience.

3. **The Importance of a Systematic Plan and an Iterative Approach.** Many TAs commented on the failure of some teams to develop a systematic and deliberate plan for defining relevance, as opposed to applying an ad hoc approach. The TAs stressed the importance of an iterative process that allowed for course correction as the TAs' view of relevance was refined through the course of the document review. The TAs identified two approaches, used by some of the teams, as particularly helpful.

The first involved the team generating and posing a sequence of questions that forced the TA to articulate increasingly precise determinations of relevance over time. Such questions helped the TA and the team to identify the outer boundary of relevance, as well as to focus on the “core” documents, as opposed to those that were only peripherally or technically relevant. The TAs felt such interactions provided them with the opportunity to reflect on and therefore to provide the teams with more well-articulated and nuanced differentiations, rather than supplying only rough, bright line rules. The TAs observed that some of the teams failed to ask questions about “grey areas.” Instead, they tended to interact in ways that produced only an undifferentiated, “black and white” understanding of relevance – often late in the process – when they would have benefited more from allowing time to raise follow-up questions designed to refine their understanding, especially in addressing “gray areas.” (The limitations of this approach were particularly apparent in requests for appeals and adjudication of human assessor determinations of relevance where “black and white” determinations made by certain teams failed to reflect further refinements the TAs had provided to other teams.) The TAs noted that some teams did use an iterative approach, but few did enough of it. The TAs tended to prefer “live” (telephone) contact with the teams over email for this kind of interaction because it provided the teams with greater exposure to the TAs thought process and more opportunity for collaboration and immediate feedback.

The second approach the TAs found useful was when teams tied questions to specific document exemplars, rather than asking for relevance determinations about whole categories of documents in the abstract. The TAs found it easier and more useful to respond to concrete examples, especially if the sample was provided in such a way as to allow the TA time to reflect in advance of having to respond, thereby permitting a more considered determination and an opportunity to formulate useful guidance for the team.

4. **Search is Challenging and Keywords Alone are Often Insufficient.** The TAs were surprised that the topics proved quite challenging for both teams and assessors, even such seemingly straightforward topics as “football” and “shredding.” For some of the topics, this may have been due, in part, to the fact that they yielded relatively few relevant documents compared to the total number of documents in the collection. Moreover, teams working on the same topic did not seem to identify the same core of relevant documents; a fair number of documents were identified as relevant by one team, but not by another. The TAs would have liked a better understanding of the process employed by each team to identify relevant documents and expressed concern about their ability to

provide an accurate description of the process if they had been forced to defend it to an adversary or court. The TAs agreed that the use of keyword search terms, alone, was not well suited for identifying “all” documents referring or relating to complex topics such as those presented by the requests for production. Unsurprisingly, the TAs observed that search terms tended to result in a less focused effort that yielded many unresponsive documents. Some of the TAs believed that while the use of search terms may be a necessary or appropriate approach to narrow the document population for further review, keyword or Boolean search is not the optimal approach to define a set of documents for production. The TAs wondered whether and how the teams were employing “concept search” methods and other advanced technologies that are readily available in the market.

5. **Concerns About the Quality of Human Review.** Based on the volume and results of the appeals and adjudication process, some of the TAs were less confident than they were before about the quality of human review. The responsiveness determinations rendered by the human assessors included many flagrant errors, suggesting that human review may be more flawed than the legal profession currently understands and acknowledges. Although human review for relevance remains the “gold standard” by which all other methods are judged, the number and nature of errors by human assessors raised questions about whether human reviewers are capable of consistently and accurately identifying relevant documents. The TAs flagged the importance of high quality, ongoing training and supervision of reviewers, and the implementation of quality control measures. The TAs suggested that the accuracy of human review was something that the TREC legal Track should consider examining in the future.
6. **The Goal of the Exercise.** As many of the observations above reflect, the TAs’ experiences raise the question of whether the teams and the TAs shared the same understanding of the goal of the exercise. The TAs were expecting the exercise to replicate the experience of responding to a discovery request in the context of civil litigation. They expected the teams to simulate an actual service provider focused on finding all of the documents the TA would need to identify to respond to the request for production. Some teams, however, were focused on understanding how a single technique could contribute to meeting e-discovery needs; they understood that the technique would not be sufficient to fulfill a production obligation.

The TAs uniformly reported that their participation in the 2009 TREC Legal Track was a valuable professional experience that provided them with insight into the search and document review processes. They strongly encourage others – both practitioners and service providers, as well as academics – to support and contribute to the TREC legal Track by participating in future exercises, whether as draftspersons, participating teams, TAs, volunteer assessors, or by analyzing its published results and suggesting other useful lines of inquiry.

Prepared by Maura R. Grossman on behalf of herself and:

Art Bieser  
Christopher H. Boehning  
Michael Roman Geske

Howard J. C. Nicols  
David Stanton  
K. Krasnow Waterman