

TREC 2011 Legal Track – Learning Task Guidelines

(June 27, 2011)

[\[Click here for a web version of this document.\]](#)

Each year since its inception in 2006, the TREC Legal Track has included one or more tasks that model the process of identifying documents responsive to requests for production that are typical in civil litigation.

The TREC 2011 Legal Track will pose a single task (hereinafter, the “TREC 2011 Legal Track Learning Task,” or the “2011 Task”), which will require participating teams to evaluate each of approximately 670,000 documents for responsiveness to one or more requests for production. The 2011 Task will most closely resemble the [TREC 2010 Legal Track Learning Task](#) (for reference, please see the 2010 [guidelines](#), [dataset](#), [draft overview paper](#), and [results/toolkit](#)).

The 2011 Task will use exactly the same participation categories, dataset, submission format, and evaluation measures as the 2010 Learning Task.

The 2011 Task will use [three new requests for production \(topics\)](#), so that all participating teams will start with "zero knowledge" as to the responsiveness of particular documents, beyond what may be inferred from the wording of the requests for production, and the contents of the documents. Our expectation is that all participating teams will complete all three topics, however, teams lacking the resources to complete all three topics may submit results for one or more topics of their choice.

For each topic, a Topic Authority (“TA”) has been assigned. The TA will (i) interpret the production request, prepare a set of “coding guidelines;” (ii) conduct a “kick-off” conference call to explain the topic to interested teams (participation in this call is strongly suggested); and (iii) provide responsiveness determinations as described below. The topic authorities are:

- Topic 401: Kevin F. Brady, Connolly Bove Lodge & Hutz LLP.
- Topic 402: Brendan M. Schulman, Kramer Levin Naftalis & Frankel LLP.
- Topic 403: Robert Singleton, Squire, Sanders & Dempsey (US) LLP.

For each topic, each participating team may request from the Topic Authority an authoritative determination of responsiveness for up to 1,000 documents in the collection. The Topic Authority for each topic is a senior litigator who has been designated by TREC to interpret the request for production (topic) and to determine the responsiveness of documents according to that interpretation.

Participating teams will request and receive responsiveness determinations using a [web interface](#). Teams may request determinations at any time, although there will be a limit of 100 documents (or determinations) that may be requested of the TA per topic, per team, in any given 48-hour period. The timeliness of responses will be determined by the availability and capacity of the Topic Authority, however, our goal is to have TAs provide responses within 48 hours, for up to 100 documents, per topic, per team. In addition, teams should not expect to receive determinations for more than 100 documents per topic in the week preceding the deadline for a submission, and should therefore plan their requests to the TA accordingly.

Teams will be required to submit interim results using the [NIST submission form](#). These interim submissions will be evaluated along with the final submission. For each topic it is undertaking, each team must submit results according to the following schedule:

- prior to requesting a determination of responsiveness on any document(s) from the TA;
- after the 100th determination and before the 101st ;
- after the 300th determination and before the 301st ;
- after all determinations requested by the team, or the 1000th determination.

Following the final submission deadline (August 28), all determinations requested by all participating teams will be released. During the subsequent week, each team is required to submit a final “mop up” run that makes use of all of all available determinations. These mop-up runs will be evaluated separately.

Participation Categories

Participating teams may choose one of two categories:

- Automatic: no human (other than the TREC Topic Authority) will view the topics or any of the documents in preparing the submission.
- Technology-Assisted: any combination of human input and technology may be used in preparing the submission.

Regardless of the participation category, a brief outline of the method must be provided when the results are submitted.

Submissions

The results for one or more topics must be encoded in a text file according to the standard TREC format, where each line contains:

- *requestid Q0 docid rank estP runid*

requestid is one of 401, 402 or 403, identifying the production request. Q0 is a historical artifact of the TREC format. *docid* is a TREC-assigned document identifier. *rank* is the ranking of the document by *estP*, where 1 is the most likely relevant document for the request. *estP* is a probability estimate between 0.0 and 1.0. *runid* is a unique identifier for the submission, formed by joining

- a sequence of 3 characters identifying the team (composed by the participating team)
- a sequence of 3 characters identifying the method (composed by the participating team)
- Capital “A” if the method is fully automated; capital “T” if both automated and manual methods are used.
- 1 for the first interim submission (prior to any responsiveness determinations); 2 for the second interim submission (prior to the 101st responsive determination); 3 for the third interim submission (prior to the 301st responsiveness determination); F for the final submission (after the final responsiveness determination requested by the participating team, or the 1000th responsiveness determination); and M for the mop-up submission.

Participating teams may submit up to three runs, which may employ different methods. However, teams may receive only one quota of responsiveness determinations per topic, and are required to submit only one set of interim results.

Evaluation

Submissions will be evaluated according to two criteria:

- How well participating teams rank (prioritize) documents from most to least likely to be responsive. The evaluation measures will be: precision, recall and F1 at various cutoffs; area under the receiver operating characteristic curve (AUC).
- How well participating teams estimate precision, recall and F1 at various cutoffs.

Evaluation measures will be computed using the [TREC 2010 Legal Learning Task evaluation toolkit](#).

Timeline

- Past Due!! -- Register for TREC, see [TREC 2011 Call for Participation](#)
- June 27 -- [Topics released](#); TA responsiveness [determination request web site live](#); [interim submission web site live](#) (done)
- July 7 – Coding guidelines (prepared by the Topic Authorities) released; introductory “kick-off” conference calls with the TAs scheduled for each topic
- August 28 -- Final submission deadline
- August 29 – All TA responsiveness determinations released
- September 5 – Mop-up submissions due
- Mid- September -- TREC speaker proposals due
- September 30 -- Results available
- Mid-October -- TREC Notebook papers due
- November 15-18 -- TREC Workshop