Interactive Task Guidelines
TREC 2009 Legal Track

Abstract

In 2009, the TREC Legal Track will again be featuring an Interactive Task (along with a Batch Task, for more on which see the Legal Track website [1]). This document contains guidelines for this year's Interactive Task, with particular focus on aspects of task design that are new in 2009; the document also covers specific steps those interested should take to register for and begin the task.

1 Introduction

In 2008, the TREC Legal Track, seeking to develop an exercise that modeled more completely and accurately the task of reviewing documents for responsiveness to a request for production in civil litigation, introduced a completely redesigned Interactive Task (see 2008 Interactive Task Guidelines [2]). The task saw participation from four teams (two from academia, two from the e-discovery industry) and produced interesting results, both with regard to the effectiveness of the approaches evaluated and with regard to the evaluation design itself (see Overview of the TREC 2008 Legal Track [4]). In 2009, the Legal Track will build on these results with a second running of the Interactive Task.

This document reviews key features of the design of the Interactive Task. The focus of these guidelines is on features new to the 2009 design; for additional background on the motivation and reasoning behind the task, readers are encouraged to review the 2008 guidelines [2].

We begin with a review of the main elements of the Interactive Task (Section 2), then turn to a discussion of modifications we propose to introduce for the 2009 running of the task (Section 3), and conclude with a summary of the specific parameters that will define the 2009 Interactive Task (Section 4).

2 Overview of Task

There are a number of uses to which an attorney may put a document collection (whether that of the party the attorney represents or that produced by an opposing party) in the course of a lawsuit or investigation. The attorney may, for example, in building a theory of the case, want to conduct exploratory probes of the collection to test whether particular hypotheses are supported by the data; the attorney may wish to gather together the documents needed to prepare a witness for deposition; the attorney may look to the collection to fill a very specific gap in an otherwise well-developed story; and so on. To each of these uses corresponds a distinct retrieval task, distinct both in terms of the objectives and in terms of the conditions in which the retrieval effort is conducted. The specific task that the Interactive Task models is that of reviewing documents for responsiveness to a request for production that has been filed during the discovery phase of a lawsuit.

In order to model the conditions and objectives of that task, and to evaluate how effective various information-retrieval methods are at performing the task, the Legal Track, in 2008, introduced a number of new design features to the Interactive Task. Some of the features pertained to the conditions and objectives of the task, others to the protocol whereby submissions would be evaluated. The most salient features of the Interactive Task are the following.

2.1 Task Conditions and Objectives

With regard to conditions and goals, the distinctive features of the Interactive Task are: (i) the role of the Topic Authority, (ii) the provision for interaction with the Topic Authority, and (iii) the definition of the task objective.

- **Topic Authority.** The Topic Authority plays the role of a senior litigator who is charged with overseeing a client's response to a request for production and who, in that capacity, must certify to the court that their client's response to the request is accurate and complete (commensurate with a reasonable and good-faith effort). In keeping with that role, it is the Topic Authority who, taking into account considerations of genuine subject-matter relevance as well as pragmatic considerations of legal strategy and tactics,[1] holds ultimate responsibility for deciding what is and is not relevant to a target topic (or, in real-world terms, what is and is not responsive to a document request). The Topic Authority's role, then, is to be the source for the definitive conception of responsiveness that each participating team, in the role of a hired cohort of manual reviewers or of a vendor of document-retrieval services, will be asked to replicate across the full document collection. Each topic has a single Topic Authority, and each Topic Authority has responsibility for a single topic.

  In performing this role, the Topic Authority has three chief responsibilities. The first is topic clarification; the Topic Authority, in accordance with procedures described below, acts as a resource to which teams can turn in order to clarify the scope and intent of a target topic. The second is review oversight. In order to be able to obtain valid measures of effectiveness, it is essential that the samples reviewed for purposes of evaluation be assessed in accordance with the Topic Authority's conception of relevance; it is the role of the Topic Authority to provide assessors with guidance as to what is and is not relevant. The third is final adjudication. As discussed below, as a further corrective on sample assessments, teams are given the opportunity to appeal sample assessments they believe have been made in error; it is the role of the Topic Authority to render a final judgment on all such appeals.

- **Interaction with the Topic Authority.** If it is the Topic Authority who defines the target (i.e., who determines what should and should not be considered relevant to a topic), it is essential that provision be made for teams to be able to interact with the Topic Authority in order to gain a better understanding of the Topic Authority's conception of relevance. In the Interactive task, this provision takes the following form. Each team can ask, for each topic for which it plans to submit results, for up to 10 hours of a Topic Authority's time for purposes of clarifying a topic. A team can call upon a Topic Authority at any point in the exercise, from the kickoff of the task to the deadline for the submission of results; in the final weeks of the exercise, however, the amount of time a team may ask of a Topic Authority may be restricted (in order to avoid an undue call on the Topic Authority's time in the final weeks of the exercise). How a team makes use of the Topic Authority's time is largely unrestricted: a team can ask the Topic Authority to pass judgment on example documents; a team can submit questions to the Topic Authority by email; a team can arrange for conference calls to discuss aspects of the topic. The one constraint (apart from the 10-hour maximum) we place on communication between the teams and their designated Topic Authorities is introduced in order to minimize the sharing of information developed by one team with another; while we instruct the Topic Authorities to be free in sharing the information they have about their topics, we also ask that they avoid volunteering to one team specific information that was developed only in the course of interaction with another team.

- **Task Objective.** When an attorney vouches for the validity of a document production, he or she is vouching for the accuracy of a binary classification of the document population implicated by the litigation, a classification into the subset of the population that is responsive to the requests for production and the subset that is not. When an e-discovery firm supports an attorney in this effort, it must make a similar responsiveness determination. The Interactive task, modeling this requirement,

---

[1]E.g., to what extent a broad interpretation of responsiveness, one that reduces the likelihood of a challenge for underproduction, will serve the client better than a narrow interpretation of responsiveness, one that reduces the likelihood of producing potentially damaging material that arguably could have been held back.

specifies that each team's final deliverable be a binary classification of the full population for relevance to each target topic. Teams are of course free to use relevance ranking as a means of arriving at their result sets, but the final deliverable is a single binary classification (relevant/not relevant) of the full population of documents.

2.2   Task Evaluation

With regard to task evaluation, key features of the Interactive Task are: (i) the metrics by which each submission's effectiveness is measured; (ii) the use of large samples to estimate performance on each topic; and (iii) the inclusion of an appeals mechanism as a corrective on first-pass sample assessments.

- **Metrics.** The metrics by which results in the Interactive Task are evaluated are recall, precision, and, as a summary measure of effectiveness, F_1.

- **Sample Size & Composition.** The Interactive Task, like the real-world task it models, calls for a significant commitment of time, both from the Topic Authorities, who define the sets of documents targeted for retrieval, and from participating teams, who endeavor to retrieve the documents so targeted. Because of the commitment required for each topic, the Interactive Task features a relatively small number of topics; in 2008, the task featured three topics, and, in 2009, the task will likely feature a similar number (four to seven, depending on the number of participating teams and on the availability of Topic Authorities). The smaller number of topics allows us to concentrate a greater amount of available assessor resources on each topic and so to construct larger and richer evaluation samples. These larger samples enable us both to obtain more precise estimates of performance metrics and to explore much more deeply the parts of the document population that no participant found responsive (in 2008, aggregating across all three topics, a total of 4,625 documents were drawn from parts of the population no participant submitted as responsive).

  Samples are composed with an eye to bringing into view a substantial number of each team's positive and negative assessments. For each topic, the population is stratified on the basis of participant submissions, with each stratum being defined for one of the possible submission combinations (one stratum containing the documents all teams counted as responsive, another containing the documents no team considered responsive, and other strata for each of the possible subsets representing conflicting participant submissions). In constructing the samples, strata are represented largely in proportion to their size in the full population, although very large strata (e.g., the "All-Negative" stratum), though certainly well-represented in the sample, are under-represented relative to full-population proportions so as to allow us to bring into view some of the smaller strata. A more complete description of the sampling design used in the Interactive Task can be found in the 2008 Guidelines [2]; for discussion of the estimation procedures followed when using this design, see the appendix to the Overview of the 2008 Legal Track [4].

- **Appeal & Adjudication.** If the goal of participants in the Interactive Task is to replicate the Topic Authority's conception of responsiveness across the test population, it is essential that the samples used to evaluate how well participants have done in meeting that goal be assessed in accordance with the Topic Authority's conception of responsiveness. To that end, a number of measures are taken to ensure that first-pass sample assessments are aligned with the Topic Authority's criteria for responsiveness (e.g., the preparation of topic-specific guidelines, provision for on-going interaction between assessors and Topic Authorities, and so on).

  Nevertheless, while the guidelines and the opportunity for further clarification can be expected to reduce the scope for assessment error, they should not be expected to eliminate it altogether. In recognition of the fact that some assessment error will likely remain at the conclusion of the first-pass review, the Interactive Task makes provision for an appeal and adjudication mechanism as a second corrective on sample assessments. Under this provision, teams, at the conclusion of the first-pass review, are given access to all sample assessments so-far entered. After reviewing the assessments,

teams are invited to appeal any assessments they believe are inconsistent with guidance they received from the Topic Authority during the topic-clarification phase of the exercise. The Topic Authority then renders a final judgment on all appealed assessments. The Topic Authority is the final arbiter and there is no second round of appeal. These final, post-adjudication, assessments are the basis for estimating all task metrics.

The features so-far described are the key features that defined the 2008 Interactive Task and are features that will be carried forward in the 2009 running of the task; as already noted, additional background on these aspects of the task can be found in both the 2008 Guidelines [2] and the 2008 track overview [4]. In carrying out the task in 2008, however, we did identify some ways in which the task might be modified in order to be more informative or more efficient; to these modifications we now turn.

3   New to the 2009 Task

The defining features of the Interactive Task will remain largely unchanged in 2009. We do, however, believe that some modifications to the task will be helpful; these are being made in the interest of (i) making the task more accessible to new participants; (ii) making the task more efficient and scalable; and (iii) enabling the task to gather additional information on what methods are effective at performing the task. The modifications are the following.

- **Background Reading.** The task of retrieving documents responsive to a request for production, the task modeled by the Interactive Task, presents a number of distinct challenges and opportunities. The 2008 Topic Authorities, reflecting on their interactions with participants in last year's exercise, found that some teams, new to the world of e-discovery, would have been well-served by a general introduction to the goals and conditions that characterize that world. Following up on this suggestion, we will provide participants with a brief list of background materials that may help to introduce them to the real-world circumstances that are both the motivation and the anchor for the task.

- **Topic-Specific Kick-Off Calls.** In 2008, we held a general kick-off call for the task, in which key features of the design of the task were reviewed and participants were invited to ask any initial questions they had about the task as they got under way with their efforts. We did not, however, make provision for any topic-specific kick-off calls, calls at which all participants who had signed up for a given topic could receive an initial orientation to the topic from the topic's designated Topic Authority. The 2008 Topic Authorities, reflecting on the task, found that such a call could have proven helpful, both to head-off any initial misconceptions about a topic and to establish a common starting point for participants.

  Accordingly, in 2009, we propose to supplement the general task kick-off call with a topic-specific call for each topic. Each topic-specific orientation call will be approximately one hour in length and will count, for any team that chooses to join the call, against the team's allotted 10 hours of time with the Topic Authority. Participation in the call is not required but is encouraged.

- **Sampling.** In 2008, in constructing evaluation samples, we defined a distinct stratum for each possible combination of participant assessments and drew at least some documents from every stratum so defined. In 2009, if a greater number of participants subscribe to each topic, it may not be possible to sample from every logically-possible stratum and still remain within a reasonable budget of volunteer assessments. In such an eventuality, we will simplify the stratification scheme by selectively merging some of the initially-defined strata that have smaller numbers of documents. While some granular control may be lost through such merging of strata, the calculation of basic measures of effectiveness (recall, precision, F_1) will not be hindered.

- **Adjudication.** In 2008, when teams submitted their appeals, the documents in which they recorded and explained their appeals were passed more or less directly to the Topic Authority for final adjudi-

cation of the assessments in question. The appeals from the various teams were not consolidated into a single document; nor were the appeals anonymized prior to the Topic Authority's adjudication.

In 2009, prior to transmittal to the Topic Authority, all appeals documentation for a given topic will be consolidated and anonymized; this step will both serve the interest of efficiency and preclude the possibility of bias (however unintentional) in rendering a final judgment on appeals.

- **Resource Usage.** In 2008, the Interactive Task attracted participation from a range of different teams, some from academia and others from industry. While such a mix suits the exercise very well, it is also true that different participants will have different amounts of resources they are able to commit to the task; an academic team, for example, may well not have access to the resources that an industry participant does. We would like to be able to account for such differences in evaluating results.

  While we want to be able to account for such differences, we also recognize that industry participants, for sound business reasons, may not wish to make full disclosure of all aspects of the process by which they prepared their submissions or the amount of resources used to do so. Such information may be proprietary and may pertain to sensitive issues of time and cost.

  In order to capture the broad distinctions we are interested in making (e.g., the distinction between an approach that called upon a lot of resources and an approach that called upon a little), we encourage that each participant, when submitting results for a topic, report the number of hours expended in preparing those results. The number of hours can be reported in as specific (x hours) or as general ($x \pm y$ hours) a manner as the participant deems appropriate. The reporting of this information is encouraged but not required.

  We believe that, by gathering this information, we will be able to account for broad disparities in resources, without adding an unnecessary disincentive for members of the e-discovery industry to participate in the task. Of course, any participant may include more detailed information on resource usage in its paper.

- **Predicted Levels of Recall & Precision.** Blair and Maron's 1985 study famously found that expected levels of recall could diverge significantly from actual levels [3]. We believe that, also in the current study, a comparison of expected to actual levels of recall and precision could yield interesting results. In order to gather data needed for a comparison, we will give participants, when submitting results for a topic, the option of reporting predicted levels of recall and precision achieved on that topic; these predicted levels can then be compared to the estimates of actual levels we obtain through the evaluation protocol. Supplying predicted levels of recall and precision is optional and is not required for any team.

4   Summary of Task Specifics

We conclude the 2009 Interactive Guidelines with a review of specific requirements for registering and executing the task. This section covers, in particular, the following: (i) task prerequisites and registration; (ii) the test collection; (iii) the unit of measurement; (iv) target topics; (v) submission of results; (vi) task schedule; (vii) reporting of results; and (viii) additional information.

4.1   Prerequisites & Registration

We welcome new and returning participants; prior participation in TREC or in the Legal Track is not a prerequisite. By way of prerequisites, all that we ask is that those considering participation carefully review these task guidelines (along with supporting materials, such as the 2008 Guidelines [2] and the 2008 Legal Track Overview [4]) in order to understand the task requirements and expectations. If there are any questions regarding the task design, please do not hesitate to contact Bruce Hedin (email address given below).

Registration for the Interactive Task is a matter of three steps.

1. Register your intent to participate in TREC 2009; do so by entering the information requested on the TREC application to participate at: http://ir.nist.gov/trecsubmit/application.html. Once your registration information has been processed, you will be added to the mailing list for TREC participants and will receive a "Welcome" message containing other details about participation in TREC.

2. Join the Legal Track mailing list, if you have not already done so. Contact oard (at) umd.edu to be added to the list.

3. Send an email to the Track Coordinators indicating that you will be participating in the Interactive Task. Once topics are available (for more on target topics, see Section 4.4), you will also be asked to send Bruce Hedin a note specifying (i) the number of topics for which you would like to submit results and (ii) a list of your preferred topics (together with some alternatives) in order of preference. We will endeavor to give all teams their preferred topics, but, in the interest of load-balancing among the Topic Authorities, we may have to ask some teams to take a topic other than their first preference. Email addresses for the Track Coordinators are as follows:

Jason R. Baron	jason.baron (at) nara.gov;
Bruce Hedin	bhedin (at) h5.com;
Douglas W. Oard	oard (at) umd.edu;
Stephen Tomlinson	stephent (at) magma.ca.

In addition to the Interactive Task, the 2009 Legal Track also features a Batch Task; for additional information on this task, see postings on the Legal Track website [1].

4.2 Test Collection

The test collection for the 2009 Interactive Task will be a collection of email messages, with attachments, produced by Enron in response to requests by the Federal Energy Regulatory Commission ("FERC").

There is no single "Enron Collection." Rather, there are many collections of Enron email. By far the most widely used such collection was assembled by MIT, SRI and CMU soon after the emails became available from FERC. Because that collection is now available from CMU, we refer to it as the CMU Enron collection.

The CMU Enron collection has two major deficiencies from the perspective of the TREC Legal Track. First, it contains about half as many unique emails (after deduplication) as are presently available from Aspen Systems (which distributes the collection to the public on behalf of FERC). Over the years, FERC has withdrawn some emails from the collection for various reasons, and they have released additional emails when the basis for withholding those emails expired. The net effect has been a substantial increase in the quantity of available email over the past several years. Second, for practical reasons, no attachments are included with the CMU Enron collection. The attachments are, however, available from FERC. Together, these factors made the CMU version of the Enron collection suboptimal for use in the TREC Legal Track.

The obvious alternative, having each participant obtain the collection from Aspen Systems, was also not practical for two reasons. First, the collection distributed by Aspen Systems can change over time, thus complicating future use of relevance judgments developed for the track. Second, the attachments are available from Aspen Systems only in native form; using those attachments would require that each team perform content extraction and format normalization independently. Cost may also be a factor for some track participants; Aspen Systems charges a fairly large fee (over $10,000) to cover their distribution costs.

We therefore chose to process a recent version of the collection provided by Aspen Systems to facilitate distribution and use for evaluation. To facilitate use for evaluation, all identifiable text and some metadata for messages and attachments were extracted, the result was exported using the EDRM XML interchange format, and two deduplication passes were performed. The resulting collection contains 569,034 unique messages. To facilitate distribution, files of the same type were aggregated (using the Unix tar utility) and compressed (using the Unix gzip utility). Distribution will be via FTP. Because we anticipate that there may be some need for bug fixes, at present the collection is being distributed only to Legal Track participants (and to others who are helping with data preparation). Ultimately, it is our plan to release a stable collection for broader research. The files are packaged in a way that participants who wish to work with only the

extracted text and metadata (or with any other subset of the collection that we can reasonably anticipate) can download only the parts of the collection that they need.

This version of the collection has two important known deficiencies that we hope to address in the near future. First, the collection as released does not contain .pst files (a format that some track participants have expressed interest in). Second, we do not presently have a mapping between this collection and others that participants could use to determine which messages that they may have already processed (after obtaining them from some other source) are present in, and absent from, this collection. If we are, with the help of others, able to address these issues expeditiously, we would intend to release an updated collection. It is, however, not our intent to change the content of this collection — absent the discovery of some catastrophic bug, we would intend to keep the contents of this present version (1.0) of the collection as the authoritative version for the Interactive Task of the TREC 2009 Legal Track.

In order to get the collection, please send an email request to Tamer Elsayed (telsayed@cs.umd.edu), cc to Doug Oard (oard@umd.edu) and Bruce Hedin (bhedin@h5.com). Tamer will then send you instructions for how to obtain the collection by FTP.

Teams wishing to work with the raw unprocessed collection (containing Microsoft .pst files, etc.) can order the collection on disk from Aspen Systems. Teams choosing to do so, however, should be aware that they will then be responsible for matching each message in their processed version of the collection to one of the processed versions above in order to ensure that they process exactly the same documents as other teams (FERC occasionally adds messages to, and sometimes deletes messages from, their public release).

Important Note on Privacy Protected Materials in Enron Public Data Set. As referenced above, this year the TREC Legal Track is using a version of the Enron Public Data Set for purposes of conducting the Interactive Task. The data set is being made available to participants in the TREC Legal Track behind a password-protected site. Registrants and others should be on notice that some individual documents in the collection may contain information of a privacy-protected or otherwise sensitive nature (e.g., social security numbers, personal and medical information pertaining to named individuals, etc.). In light of existing privacy laws, any individual or organization that downloads the Enron Public Data Set should use reasonable and appropriate safeguards in maintaining the confidentiality of privacy-protected information, including with respect to future publication of the results of the research.

4.3   Unit of Measurement

As has been noted, the documents in the Enron Collection consist of both parent emails and child attachments. In the period of time allocated for public comment on this year's draft Interactive Task guidelines, the members of the Legal Track community had a lively and productive discussion on the question of whether the unit for evaluation should be the "record" (that is, a parent email together with all of its attachments) or the "document" (that is, each member (parent email or attachment) of the full record considered separately).

Some participants in the discussion argued in favor of record-level measurement, grounding their arguments primarily in the fact that discovery requests generally require the production of the full record once any member of the family is found to be responsive. Other participants argued in favor of document-level measurement, grounding their arguments chiefly in that (i) document-level metrics, being more granular, are a more sensitive gauge of retrieval effectiveness, (ii) propagation of responsive status from an individual document to all members of the full record is a mechanical matter and not one of retrieval effectiveness, and (iii) there are some use cases in which an awareness of just which members of the full record were responsive would bring efficiencies.

In light of the perspectives expressed in this discussion, we have decided to report measures at both the record and the document level. Participants are free to focus their efforts on either record-level or document-level performance (or both), but they should be aware that results will be reported at both levels.

4.4 Target Topics

As in 2008, topics will take the form of document requests associated with a mock complaint. The number of topics offered will be similar to the number offered in 2008 (in the range of 4–7, depending on the number of participating teams and on the availability of Topic Authorities). How many topics a participating team takes on is largely up to the team; in order to prevent any one Topic Authority from being overburdened, however, we may have to limit the total number of topics for which a team can sign up. We will endeavor to assign each team its preferred topics; should, however, any one topic be heavily subscribed, and therefore any one Topic Authority potentially overburdened, we may have to ask a team to take an alternative topic.

The mock complaint and the associated topics will be released upon the official start of the exercise.

4.5 Submission of Results

Specifics of the procedures and format for submitting results will be provided to participants in a separate document. For present purposes, please note the following.

- Submissions must be made using Legal Track Interactive Task Results form, which will be accessible from the TREC Information for Active Participants page.

- Each team should prepare and submit a single file containing results for all topics for which a team is submitting results for evaluation.

- The submission file should list only documents found relevant to each target topic. Any document included in the submission list for a given topic will be assumed to have been found relevant to that topic; any document not included in the submission list for a topic will be assumed to have been found not relevant to the topic.

- Submissions must be made by the deadline noted below (Section 4.6); submissions made prior to the deadline will be accepted and are welcome.

- A team should submit one set of results for each of its target topics; it will be possible to obtain scores for alternative result sets once the evaluation samples have been selected, reviewed, and adjudicated.

4.6 Task Schedule

The timeline for the 2009 Interactive Task follows.

Guidelines frozen; topics released	06/19/09
Window for task execution	06/19/09 – 09/07/09
Window for task registration closes	07/02/09
Deadline for submitting results	09/07/09
First-pass review of evaluation samples	09/08/09 – 09/30/09
Preliminary (pre-adjudication) metrics released	10/02/09
Appeal & adjudication	10/02/09 – 10/22/09
Final (post-adjudication) metrics released	10/23/09
Conference papers due	10/26/09 (estimate)
TREC 2009	11/17/09 – 11/20/09
Final papers due	February 2010

4.7 Reporting of Results

As in past years, Track Coordinators will prepare a report summarizing results from the Legal Track. A preliminary draft of the report will be prepared for the November 2009 meetings in Gaithersburg; a final version of the report will be made available in early 2010.

Individual teams are also expected to prepare reports of their findings. Preliminary drafts are submitted in late October, with revised versions being posted on the TREC Web site in February 2010.

At TREC (November 17-20, 2009), all participants will be given the opportunity to present their results at the Legal Track Workshop. Participants are also encouraged to submit proposals for presentation in the Legal Track plenary session.

4.8   Additional Information

The Legal Track website at http://trec-legal.umiacs.umd.edu/ contains links to additional resources and background information. The track mailing list archives can be reached through a link from that page. For further information on the Interactive Task in particular or the Legal Track more generally, please contact one of the track coordinators:

Jason R. Baron	jason.baron (at) nara.gov;
Bruce Hedin	bhedin (at) h5.com;
Douglas W. Oard	oard (at) umd.edu;
Stephen Tomlinson	stephent (at) magma.ca.

References

[1] The Legal Track Website, 2009. Available at http://trec-legal.umiacs.umd.edu/.

[2] J. R. Baron, B. Hedin, D. W. Oard, and S. Tomlinson. Final Interactive Task Guidelines, 2008. Available at http://trec-legal.umiacs.umd.edu/2008InteractiveGuidelines.pdf.

[3] D. C. Blair and M. E. Maron. An evaluation of retrieval effectiveness for a full-text document retrieval system. *Communications of the ACM*, 28(3):289–299, March 1985.

[4] D. W. Oard, B. Hedin, S. Tomlinson, and J. R. Baron. Overview of the TREC 2008 Legal Track. In *The Seventeenth Text REtrieval Conference (TREC 2008) Proceedings*, November 2008.