TREC 2010

Independent Third-party Testing and Validation for Equivio>Relevance

equivio

zoom in. find out

OVERVIEW

With predictive coding technology approaching an adoption tipping point in the electronic discovery community and its acceptance growing among lawyers and judges, the results of the 2010 Text Retrieval Conference ("TREC 2010") Legal Track are particularly compelling in the current climate. As discovery costs skyrocket and the volume of potentially applicable information explodes, resource limitations are the hallmark of modern litigation. In fact, over 40% of the largest U.S. corporations planned to increase their spending on electronic discovery in 2011 because of this shift, reports Fulbright & Jaworski in its 7th Annual Litigation Trends Survey Report.

These concerns have prompted the government to increase its international research efforts aimed at studying information retrieval in large-scale document evaluations. Accordingly, for the past few years, the National Institute of Standards and Technology ("NIST"), an agency of the U.S. Department of Commerce, has co-sponsored an interactive document review exercise that features both academic and commercial participants interested in chemistry research, medical records, and electronic discovery, among others.

The Legal Track, divided into batch and interactive tasks, seeks to objectively model the e-discovery review process for purposes of evaluating the efficacy of various search techniques and tools. To accomplish this task, NIST openly invites organizations focused on leveraging technology to hone the accuracy of data selection by collaborating on a series of discovery-oriented activities using a standard set of records. Each participant then shares its results and contributes to a collective discussion toward future progress.

Since its inception in 2006, the TREC Legal Track has gained broad acceptance. In its open letter to "strongly encourage" the participation of e-discovery service providers, law firms and corporate counsel, the Sedona Conference expresses its full support for the joint creation of a "credible, collaborative and independent process" to assess existing and emerging search methods. Courts have similarly embraced this initiative, with Judge Grimm prominently expressing his support in Victor Stanley v. Creative Pipe, and more recently by Judge Andrew Peck in his article in Law Technology News.

THE CHALLENGE

The interactive task models the aspirations and constraints of an authentic document review project. Each team focuses on a comprehensive search for records that are responsive to a request for production in a civil lawsuit. Their goal is generally to gauge the effectiveness of digital tools that are designed for use in complex litigation.

TREC 2010 offered contributors the chance to explore 685,592 e-mails from the public Enron bankruptcy file in this effort (it provided material from similar records for TREC

2009 as well). Each team received a mock complaint that set forth the prevailing law and critical facts associated with the sample case along with a choice of issues to which it could respond. Equivio chose to use its Equivio>Relevance analytical tool for the interactive task to respond to document request 303. Request 303 required the production of all material germane to Enron's lobbying efforts.

METHODOLOGY

In order to execute this task, Equivio collaborated with a "Topic Authority" assigned by the administrators of the exercise. For each team, the Topic Authority assumed the role of a senior lawyer responsible for replying to that particular request for production and certifying its accuracy to the court pursuant to a reasonable good-faith effort.

As such, the Topic Authority was the ultimate arbiter of responsiveness and, therefore, Equivio leveraged the knowledge of the Topic Authority while an Equivio appointed attorney (aka the "expert") trained Equivio>Relevance to make similar determinations. Equivio>Relevance uses statistical and self-learning techniques to calculate graduated relevance scores for each document in the data collection.

Weighing hundreds of distinct characteristics, including keyword frequency and proximity from one vital term to another Equivio>Relevance continuously refines its relevance criteria through an iterative, self-correcting process. As the expert reviewed additional documents, the software was able to improve the accuracy of its scoring system. Once it achieved a threshold level of correctness, it ranked the entire data set and assigned each document a relevance score. Figure 1 shows a typical learning curve for training of the Equivio software to identify relevant documents. The plateau effect, visible in this learning curve from sampling batch 15 and onwards, indicates that the learning process has optimized.

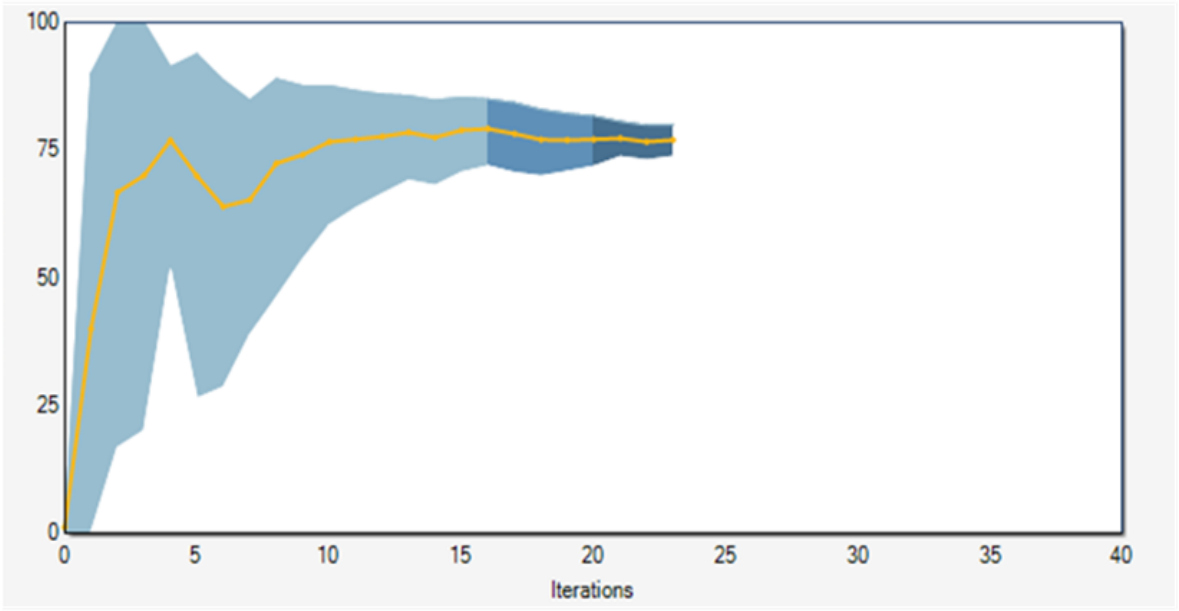

Figure 1. Equivio>Relevance learning curve

In addition to verifying relevance assessments with the Topic Authority, Equivio sought to more closely align this exercise with the approach taken in a standard law firm. It, therefore, assigned a "peer consultant" to assist in the review of documents deemed to be of questionable applicability. This peer consultant, also experienced in litigation and standard review practices, participated in all communications between Equivio and the Topic Authority, and in so doing fulfilled a role similar to a law firm associate.

PROFILE

Overall, for request 303, the Equivio expert and peer consultant reviewed 6,200 documents. This included 2,320 documents that the expert used to train the system. The training documents were reviewed over 58 training rounds (aka batches or iterations), each comprising 40 sample documents.

In addition to the training phase per se, the 6,200 documents reviewed also included documents reviewed by the team in order to familiarize themselves with the topic and the collection, as well as quality assurance to statistically verify the success of the software in identifying relevant documents.

In terms of time spent, the expert conducted the training phase over 28 hours. As is common in litigation, the expert, the peer consultant and the Equivio team spent an additional 90 hours on the case. The expert and peer consultant invested significant time prior to the commencement of the training procedure to familiarize themselves with the case and the collection. The team also spent substantial additional time preparing for, documenting and analyzing the results of its meetings with the Topic Authority. The

actual interaction with the Topic Authority required 9 hours. The remainder of the time spent included internal team discussions, designing the quality control protocols, the actual quality assurance review itself, as well as standard administrative and management tasks. Batch scoring of the population took 42 minutes.

During the quality control phase, the Topic Authority and peer consultant participated in a "Discrepancy Analysis" to verify the results of the automated review. To complete this analysis, Equivio compared a series of manual assessments on random sets of documents to the relevance designations generated by Equivio>Relevance.

RESULTS

The TREC 2010 Legal Track is based on a collaborative and collective effort to evaluate the use of advanced technology in a complex document review project. TREC has conducted tests and found that Boolean keyword searches alone locate, on average, approximately a quarter of the total number of relevant documents. In contrast, of the initial 685,592 records available for the TREC 2010 Legal Track, Equivio>Relevance identified 26,626 responsive documents, which was merely 3.9% of the total set, but 80.1% of the relevant material.

The TREC organizing committee released the following official results of Equivio's performance:

Table 1. Equivio Results for Request 303, TREC 2010

Measure	Result	Margin of Error
Recall	80.1%	+/- 6.4%
Precision	57.7%	+/- 2.0%
F1 (F-measure)	67.1%	+/- 2.6%

Recall is the percentage of relevant documents retrieved, in this case by Equivio. Precision, by way of contrast, is the percentage of retrieved documents that are relevant. As illustrated in Figure 2, recall and precision are in perpetual trade-off. High recall means that most of the relevant material has been retrieved, but that there may be a lot of non-relevant documents to review (low precision and high cost). High precision means that most of the retrieved documents are relevant, but a lot of the relevant documents may have been missed (low recall and potentially greater risk). The rules of the TREC project require teams to maximize the F measure, an evenly weighted combination of recall and precision. In live cases, greater weight is usually assigned to recall in order to retrieve more of the relevant material.

Figure 2. Review-recall trade-off

CONCLUSION

Equivio's results in the TREC 2010 Legal Track show that a legal team could potentially review less than 4% of a collection of almost 700,000 documents and retrieve 80% of the responsive items. Using a lower cut-off score, Equivio>Relevance determined that a review of 20% of the population would yield recall levels in the range of 90%. Moreover, Equivio's precision rate was 57%, meaning that more than 1 in every 2 documents submitted by Equivio was in fact responsive. These results far outstrip the historical performance of keyword matching, which, as documented in the original Blair and Maron study from 1985, and subsequently substantiated in the TREC studies, typically retrieves only 20 to 30% of the relevant documents in the collection. The results provide additional confirmation to the growing body of evidence that predictive coding, properly applied, is a valid technique for filtering document collections in the e-discovery process, and should be considered a legitimate replacement for keyword searching.

It should be added that the success of predictive coding in the TREC project not only enables more efficient culling, but also opens the way to innovative new techniques and methodologies in the e-discovery process. For example, by focusing attention on the most relevant data, predictive coding enables systematic early case assessment and the informed evaluation of case winnability. Similarly, predictive coding allows prioritization and stratification of document review, accelerating case development and reducing review cost and resources. It also enhances the quality of the review, while enabling systematic quality control protocols. Ultimately, it offers a defensible alternative to a historically flawed keyword-centric approach, while facilitating a more efficient process that optimizes the use of costly resources, a key factor in the current economic environment.

About Equivio

Equivio develops text analysis software for e-discovery. Users include the DoJ, the FTC, KPMG, Deloitte, plus hundreds of law firms and corporations. Equivio offers Zoom, an integrated web platform for analytics and predictive coding. Zoom organizes collections of documents in meaningful ways. So you can zoom right in and find out what's interesting, notable and unique. Request a demo at info@equivio.com or visit us at www.equivio.com.

Zoom in. Find out.