

# EQUIVIO AT TREC 2009 LEGAL INTERACTIVE

TAL STERENZY  
tal.sterenzy@equivio.com

## TECHNOLOGY STRATEGY

Equivio participated in two runs under the legal interactive track: topics 205 and 207. The runs utilized the Equivio>Relevance product. Equivio>Relevance is an expert-guided system which enables automated prioritization of documents and keywords.

Based on initial input from a lead attorney, Equivio>Relevance uses statistical and self-learning techniques to calculate graduated relevance scores for each document in the data collection.

Equivio>Relevance works like this:

- The software generates a sample of documents, and feeds them to an expert, that is, an attorney who is familiar with the case.
- The expert tags the sample documents as relevant or not.
- These samples are used to train the software. Based on the expert's decisions for the sample documents, the software learns how to score documents for relevance.
- In an iterative, self-correcting process, Equivio>Relevance feeds additional samples to the expert. These statistically generated samples allow Equivio>Relevance to progressively improve the accuracy of its relevance scoring.
- Once the sampling process has optimized, the software indicates that it has reached "stable" mode. "Stable" mode is an indication that additional samples are not required. At this point, the batch scoring can be invoked. In this process, the software calculates a graduated relevance score for each document.

Each sample comprises 40 documents. Typically, 25 through 50 iterations, each of 40 documents, are required in order for the software to stabilize and optimize the sampling.

Equivio>Relevance comprises three key modules – an interactive model used by the expert for tagging documents, a statistical model which serves to monitor and manage the iterative sampling process, and a text classification algorithm. The statistical model manages sample selection. The samples are fed to the expert, and the input from the samples is used to train the classifier.

In culling scenarios, one of the challenges in using the software is the conversion of the relevance score (from 0 to 100), into a discrete decision (responsive or not responsive). In situations where the relevance scores are used to prioritize review activities, as in early case assessment or detailed review, no cut-off score needs to be defined. However, this is a critical part of the culling task, as manifest in the TREC project. In order to deal with this challenge, the software uses a statistical model for the estimation of recall and precision. In order to manage the trade-off between recall and precision, the software maps the F-measure statistic across the relevance scores to identify the F-measure maximum point. This provides the user with a recommended starting point for selecting the cutoff relevance score. The cost of the consequent review effort is a function of the percentage of documents that survive culling, as calculated by the software. Using the tool's control panel and indicators, the user can then move the cutoff point to a lower relevance score (higher recall, lower precision) or a higher relevance score (lower recall, higher precision).

Another key challenge for the software is the quality of expert input. As widely recognized in research in this area, inconsistent training can impact the quality of text classification algorithms. In order to deal with this potential problem, Equivio applies near-duplicate technology. The software identifies, in real-time, near-duplicate documents that the expert has scored inconsistently. In some cases, small differences between documents may be decisive in determining a document's relevance. However, in many cases, such inconsistencies are the result of inadvertent errors. The inconsistent tags trigger a warning to the expert to check the accuracy of the tags for these documents.

## **TREC CASE STUDY**

When using Equivio>Relevance, the standard approach is to utilize one "expert" per case. This approach was also adopted in the TREC tests, with one adjustment – alongside the "expert", who was providing the formal input to the tool, we assigned a "peer consultant". The role of the peer consultant was to work with the expert on reviewing "grey area" documents, and to be available for the expert to discuss the parameters of relevance in the case. The peer consultant also participated in the calls with the Topic Authority, in order to ensure that the input from the Topic Authority was captured clearly and completely.

Both the "expert" and the "peer consultant" were US attorneys, with experience in litigation and litigation review processes.

As noted, Equivio participated in two runs – topics 205 and 207. For topic 207, which was the more straightforward of the two topics, the interactive tagging phase involved 32 rounds. This effort required 15 hours of the expert's time, plus 11 hours from the peer consultant. The interaction with the Topic Authority involved a total of 6 hours. The batch scoring of the population took 58 minutes.

In topic 205, the interactive phase involved 49 rounds. The additional effort required for topic 205 was a function of the low richness of the population, the inherent fuzziness of the topic itself, and the emergent definition of relevance parameters, as experienced by the Equivio team in the interaction with the Topic Authority. The number of the hours spent with the Topic Authority on this topic was also greater than for topic 207. In fact, due to the fuzziness of the topic, the TREC management committee decided to allocate additional Topic Authority time for this topic. A total of 11.5 hours of the Topic Authority's time was used. All in all, the time required for interactive tagging by the expert was 21 hours. The peer consultant devoted 3.5 hours, in addition to the time spent on the calls with the Topic Authority. The batch process, in which all the documents in the population are scored, was completed in 58 minutes.

These numbers are summarized in Table 1:

	<b>Topic 205</b>	<b>Topic 207</b>
<b>Number of sampling rounds</b>	49	32
<b>Total number of documents sampled</b>	1,960	1,280
<b>Hours spent by expert in interactive tagging</b>	21 hours	15 hours
<b>Hours spent by peer consultant</b>	15 hours	11 hours
<b>Hours spent from Topic Authority "bank"</b>	11.5 hours	6 hours
<b>Batch processing</b>	58 minutes	58 minutes