

Benchmarks for Accuracy

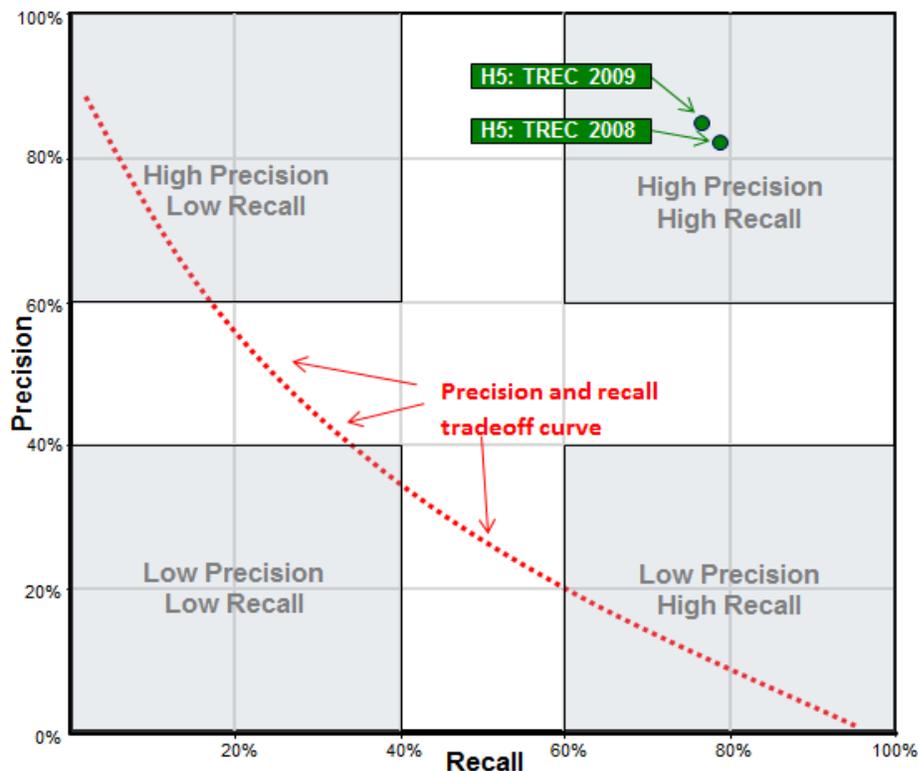
With the prominence of technology-assisted review in e-discovery, courts and practitioners alike are recognizing that the practice of “search” in e-discovery is an information retrieval problem—a scientific discipline distinct from the practice of law. Courts are also raising the bar for lawyers and litigants, requiring them to prove that the search methods they use are reasonable. It is only recently that the vendors litigating parties rely on are being held to the same standard of proof. Dozens of vendors purport to have the perfect search technology, touting high accuracy and defensibility, with nothing to back up their claims.

The Text REtrieval Conference (TREC) is an international research initiative sponsored by the National Institute of Standards and Technology (NIST), a division of the U.S. Department of Commerce. It was established in 1992 to support research for evaluation of text retrieval methodologies. The stated goal of the TREC Legal Track is “to apply objective benchmark criteria for comparing search technologies” in the context of e-discovery, providing a concrete reference by which various technologies and vendors can be independently evaluated.

The Sedona Conference’s *Open Letter to Law Firms and Companies in the Legal Tech Sector* stated “[T]he TREC Legal Track represents the emergence of an important research platform: a credible, collaborative, and independent process and protocol by which both long established and emerging search methods used for document review may be evaluated and benchmarked.” In *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251 (D. Md. 2008), Judge Paul Grimm stated: “[The TREC Legal Track] project can be expected to identify both cost effective and reliable search and information retrieval methodologies and best practice recommendations, which, if adhered to, certainly would support an argument that the party employing them performed a reasonable ESI search...”

H5's TREC Results

H5 participated in the TREC Studies evaluation in 2008 and 2009. H5's results for both years are shown by the dots in the chart below. Further information on how to read and interpret this chart is provided on the next page.



Source: See 2008 and 2009 TREC Studies at <http://trec-legal.umiacs.umd.edu/>. Note for 2008, results are shown for OCR-adjusted performance as discussed on page 36 of the 2008 TREC Overview paper.

Understanding the TREC Results Chart

Precision and **Recall** are the two fundamental review system accuracy metrics that, when taken together, best convey the overall effectiveness, or accuracy, of a document review or search process. To understand the chart above, it's important to first define Precision and Recall and why they matter.

Precision measures how many of the documents retrieved in a search are actually relevant, that is, **how much of the result set is on target**. For example, a 65 percent precision rate means that 65 percent of the documents retrieved are relevant, while 35 percent of those documents have been misidentified as relevant. Achieving high Precision means you produce only what you have to (maintain advantage) and keep costs down by reviewing only what you should (i.e. fewer non-relevant documents).

Recall measures how many of the relevant documents in a collection have actually been retrieved, that is, **how much of the target set has been found**. For example, a 40 percent recall rate means that 40 percent of all relevant documents in a collection have been found, and 60 percent have been missed. Achieving high Recall ensures you have what you need to produce (compliance) and, as important, what you need to win (you found as many relevant documents as possible).

Any document review process can achieve either high Recall or high Precision, but rarely both simultaneously. An effort to improve the performance of one factor generally causes the performance of the other to drop. This is often referred to as the "Recall-Precision tradeoff." Put simply, it is fairly easy to achieve high Precision if you are willing to risk missing most of the relevant documents (low Recall). Conversely, when most search or review methods attempt to find most of the relevant documents (high Recall) they sacrifice Precision and return large numbers of false positives. Accordingly, the key is to know what level of Precision and Recall a search or review method achieved **simultaneously**.

The chart on the previous page plots Precision on the vertical axis and Recall on the horizontal axis, with points on the chart representing the **simultaneous Precision and Recall** achieved by a review system or process. The trade off effect between Precision and Recall described above is illustrated in the downward sloping red dotted line. The goal of any search or review process is to achieve results in the shaded gray box in the upper right corner: find nearly all of the relevant documents (high Recall) and have all that you found be relevant with very few false positives (high Precision).