

An Introduction to Recall and Precision

In litigation, investigations, and preservation matters, a clearly defined method for evaluating document review performance is critical to understanding an organization's legal and financial risk.

Two measurements—recall and precision—best gauge the overall effectiveness of information retrieval systems. In the context of document review, these measurements are used to assess an organization's likelihood of missing important documents and encountering cost and time overruns during a review.

Recall

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.

Measuring recall helps organizations identify risks associated with:

- **Compliance.** Failing to identify relevant documents required for legal holds and responsive discovery requests could result in substantial penalties. High recall means the necessary documents have been found.
- **Case preparation.** Not retrieving all of the relevant documents could result in overlooking information vital to winning a case. High recall ensures a more complete set of relevant documents to review and enables more effective case preparation and strategy.

Precision

Precision measures how many of the documents retrieved 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.

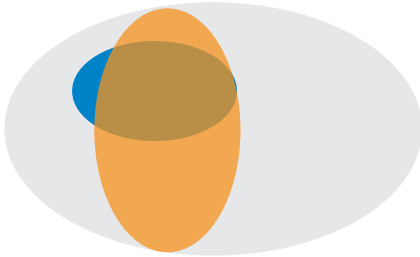
Measuring precision helps organizations identify risks associated with:

- **Expense and time.** With low precision, large numbers of non-relevant documents will need to be reviewed by attorneys. This increases costs considerably and may cause significant delays in completing the review. High precision enables more efficient review, saving legal teams substantial cost and time.
- **Preservation and production.** Many cases are lost because litigants produce documents they are not required to produce or even retain in the first place. High precision means fewer potentially damaging documents in the hands of opposing counsel.

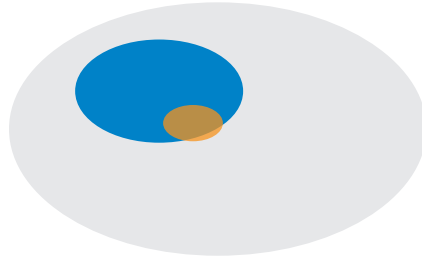
MEASURING DOCUMENT REVIEW PERFORMANCE

The Recall-Precision Tradeoff

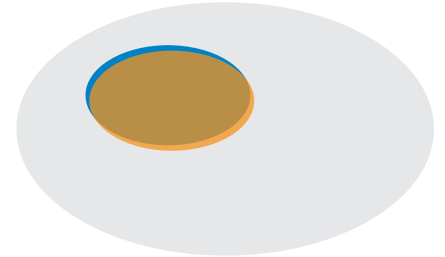
Any document review process can achieve either high recall or high precision but rarely both simultaneously. An effort to improve the performance of one factor generally causes the performance of the other to drop.



High recall with low precision is easy to achieve. You can retrieve most of the relevant documents if you cast a net wide enough. But you will also retrieve a lot of what is not relevant.



High precision with low recall is easy to achieve. If you keep your searches few and narrow, most of what you retrieve will be relevant. But many relevant documents will be missed.

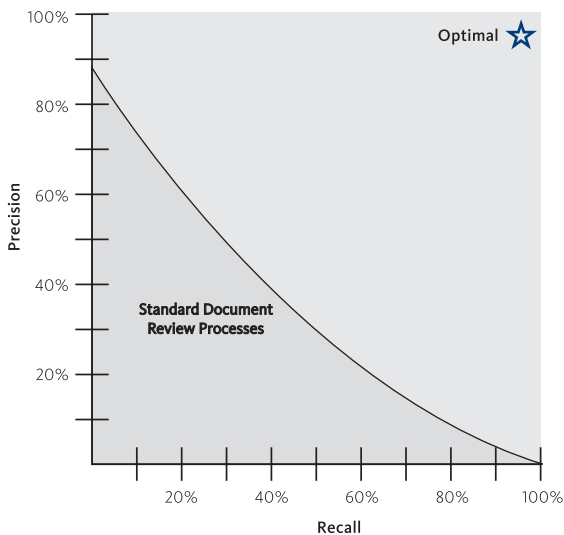


The optimal result—high recall with high precision—is difficult to achieve.

○ All documents in the collection ● Relevant documents in the collection ● Documents retrieved by the review process

The Tradeoff Curve

This curve illustrates the tradeoff between recall and precision.* Standard document review processes—whether manual or assisted by search tools—typically achieve results that fall along this curve.



* The data on best performance benchmarks illustrated on this curve derives from studies conducted by the Text REtrieval Conference (TREC), an initiative of the National Institute of Standards and Technology (NIST) and the U.S. Department of Defense.

H5 Outperforms the Curve

H5's document review significantly outperforms manual review and search-assisted review. Finding just the necessary information and little else is no easy task, but H5 does it through a combination of technology and human expertise. Our high levels of recall and precision have been verified by rigorous evaluation protocols and third-party assessments.

With measurably more accurate results, companies and their outside counsel can substantially reduce the legal and financial risks associated with document review.

Ask us to prove it.

About H5

H5 is an information retrieval firm that provides specialized search and analysis services to law firms and corporations. We advise on and execute discovery and document review strategies for litigation, investigations, preservation, and litigation readiness. Through a combination of advanced technologies and expertise in linguistics, statistics, computer science, law, and process design, we automate the assessment of documents for responsiveness or relevance to our clients' needs. The results we deliver are unparalleled in accuracy, speed, and cost-effectiveness, as validated by independent industry and academic assessments. Over the past decade, we have helped our clients realize billions of dollars in favorable outcomes in high-profile litigation and enterprise-wide information management.

© 2008 H5. All rights reserved.