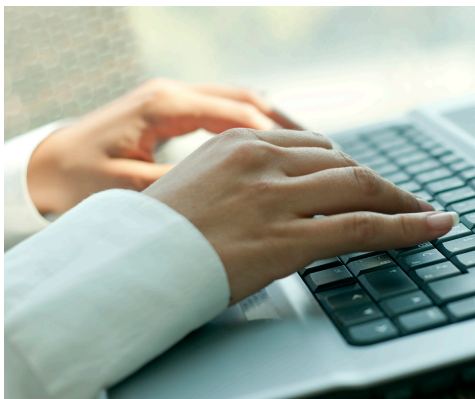


Global 50 Corporation Speeds Discovery with CategoriX

Automated document classification promotes accuracy



Over 67% savings were achieved when all two million documents were categorized for responsiveness with CategoriX classifying 93% of the documents.

The Situation

In 2009, a U.S. regulatory agency issued a subpoena to a Global 50 corporation resulting in a need to review millions of documents in just three months. In most cases, this request would have required hiring a sizable team of attorneys to review the documents. The costs associated with staffing the review would have been significant.

Since aggressively staffing to meet the deadline was not ideal, the corporation looked at other options, including advanced technologies to automate parts of the review process or reduce the document population. One option included clustering. Clustering technology offers value in certain situations, but did not support the classification capabilities the company truly needed for this particular production.

The Solution

Having used Xerox® Litigation Services' OmniX™ review platform for more than a year on other matters, the corporation turned to CategoriX, Xerox's patented automated document classification technology.

CategoriX is a text-based generative modeling technique, based on the widely accepted probabilistic latent semantic analysis (PLSA). It is Xerox® Litigation Services' approach to human-assisted computerized document review.

CategoriX was quickly approved for use by the regulatory agency, and provided greater quality control, consistency and speed than a standard manual review. Moreover, upon making the decision to use CategoriX, the company was confident in Xerox® Litigation Services' technology, high level of support and ability to work collaboratively with clients. Xerox's classification experts, along with the attorney team, drive the technology to create a reasonable and defensible process.

Confidence in methodology

A CategoriX-based review leverages a relatively small amount of attorney input to programmatically code the remaining document population. The accuracy of the review is measured by recall—a metric indicating how successfully the system captures all the relevant documents—and precision—a metric indicating how successfully the system captures only the relevant documents. Xerox® Litigation Services uses these proven information retrieval metrics to evaluate CategoriX's results, enhancing defensibility. For this agency request, the corporation emphasized recall, placing a premium on capturing all the relevant documents.

The Results

In just 54 days, using 10 attorneys, all necessary manual review input was completed—this equaled 7% of the two million document population; CategoriX automatically classified the remaining 93%. A traditional manual review of all two million documents would have taken 425 days using the same 10 attorneys, based on their average review rates from the CategoriX project. CategoriX resulted in savings of approximately 371 days using 10 attorneys or the time of an additional 68 attorneys to complete the review in the 54 days—a total of 67% cost savings for the client. In addition, CategoriX was able to achieve both high precision, 92.3%, and exceptionally high recall, 99.8%—these results are particularly impressive given the client's emphasis on recall.

Upon completion of the project the client received quantified data demonstrating the accuracy of the CategoriX results. If required, this information could have been used to defend the methodology by which the documents were produced to the agency. Most importantly, it gave the client confidence in the quality of their document review.



Save on costs



Increase speed



Minimize the number of documents for review



Enhance accuracy and consistency

Case Study Snapshot

The Situation

- Subpoena from a regulatory agency would require the manual review of two million documents in three months

The Solution

- CategoriX—a solution that combines human-assisted computerized review and intense quality control measures
- Text-based generative modeling approach relies on statistical measurements to validate accuracy

The Results

- Completed manual review and QC in 54 days
- Saved 67% in costs, which include time and resources
- Detailed results ensures methodology can be defended, and gave corporation confidence in solution
- Achieved 92.3% precision and 99.8% recall

Xerox® Litigation Services:
Voted Best eDiscovery Provider
Two Years in a Row in the 2010 and 2011
New York Law Journal Rankings

About Xerox® Litigation Services. Voted “Best eDiscovery Provider” two years in a row in the 2010 and 2011 *New York Law Journal* rankings, Xerox® Litigation Services is one of the world's largest providers of e-discovery technologies and services. With more than two billion pages of hosted data, tens of thousands of client users and over 300 employees, Xerox® Litigation Services is the trusted partner of global corporations and their outside counsel, streamlining the e-discovery process from collection through production. With our OmniX™ review platform, automated document classification technology, CategoriX and a wide range of e-discovery consulting services, Xerox® Litigation Services efficiently and effectively manages the most complex, voluminous discovery projects from beginning to end.

For more information on Xerox® Litigation Services visit www.xerox-xls.com, call 877.273.3887 or e-mail info@xls.xerox.com.

