

Utilizing Search Expertise to Reduce the Costs and Risks of Document Review

Practical Strategies to Optimize Search Term Culling with Sampling and Iterative Development

Contents

- 2 Executive Summary
- 2 Do You Know the Real Costs and Risks of Keyword Search?
- 2 The Evolving Practice of Keyword Searching
- 2 The Bench Says: Search is a Science
- 3 Practical Steps to Develop a Defensible, Iterative Approach to Search Term Culling
- 3 Key Considerations
- 4 Conclusion

July 2011

Author:

Amanda Jones

Senior Classification Analyst at
Xerox Litigation Services

Executive Summary

Legal teams are adopting new best practices to cull document volumes in preparation for document review. These approaches utilize sampling and iterative development, drawing upon technical expertise required by the courts. This paper discusses how leveraging an iterative approach that is driven by technical expertise including linguistics and statistics, informed and validated by attorneys, and rigorously tested and documented, can defensibly eliminate significant volumes of non-relevant data without sacrificing relevant data – and reduce document hosting and review costs.

Do You Know the Real Costs and Risks of Keyword Search?

Exactly how prevalent is non-relevant material in your review population? Is your case team spending the majority of its time reviewing non-relevant documents that were not culled out of the review population? Have potentially relevant documents been excluded from your review? If you are unable to answer these questions, you may be spending excessive time, money and resources on document review and exposing your organization to unnecessary risk.

So how do you effectively cull out non-relevant data to create a rich, relevant population for review? And how do you do this in a way that balances your organization's increasing demand for efficiency with your responsibility to ensure a defensible e-discovery process?

The Evolving Practice of Keyword Searching

The practice of keyword searching as part of e-discovery is evolving rapidly. Legal teams and e-discovery vendors now have access to a wide array of advanced document analysis tools that enable them to execute sophisticated culling strategies that incorporate custodian, date, file type, and keyword criteria, along with many others. These advanced culling strategies are designed to reduce hosting and review costs as much as possible by narrowing down the population to just those potentially relevant documents that actually merit review.

Unfortunately, the keyword search component of advanced culling strategies remains a source of unnecessary expense and unforeseen risk. Keyword missteps frequently lead to simultaneous over-inclusiveness – resulting in more time and money spent reviewing off-topic data – and under-inclusiveness, at the risk of missing critical information. Errors of these types are common because it is difficult to obtain a solid understanding of the linguistic content of a review population during the earliest stages of a case – but it is at that time when most culling efforts are undertaken.

Case teams are learning, though, that these challenges can be overcome by introducing sampling, iteration and linguistic analysis into the keyword development process. Investing a small amount of time upfront in a targeted analysis of keyword performance can lead to dramatic downstream cost and time savings while enhancing the defensibility of the search process.

The Bench Says: Search is a Science

Keyword search – once readily accepted by the courts and the legal community – is increasingly subject to judicial scrutiny. A few well-known cases have brought these issues to the forefront:

- In *Victor Stanley, Inc. v. Creative Pipe, Inc.*, Magistrate Judge Paul Grimm held that defendants waived privilege because they were unable to articulate which keywords were chosen, how they were applied or the qualifications of the individuals who selected them.
- In *U.S. v. O'Keefe*, Judge Facciola asserted that search involves “the interplay...of...computer technology, statistics and linguistics.”
- Judge Scheindlin, in *Pension Comm. of the U. of Montreal Pension Plan v. Bank of America Sec.*, said witnesses should be prepared to testify on “how the search was conducted, who was asked to search...and the extent of any supervision,” and “failure to assess the accuracy and validity of selected search terms... [may constitute]...negligence.”

Together, these decisions provide critical guidance on how search should be conducted to successfully defend against challenges.

Why Search Expertise is Critical

1. **Minimize hosting and review costs.**
Introducing experts early in the process can result in a deeper culling – reducing hosting costs and the time and cost spent by attorneys reviewing non-relevant information.
2. **Review a smaller, richer set of documents.**
Well-designed search terms, built with linguistic insights and thoroughly tested for precision and comprehensiveness, ensure rich and focused document collections – leading to a more expedient and informative path to case strategy and production.
3. **Ensure a better search term negotiation outcome with opposing counsel.**
Statistical data and firsthand knowledge of the quality of your review population will position you for a better outcome in Rule 26(f) negotiations with opposing counsel.
4. **Focus on defensibility throughout the process.**
“Right size” document review and efficiently fulfill legal obligations for due diligence by utilizing: a principled and well-documented approach; technical expertise called for by the courts; and statistically valid measurement of results.
5. **Focus on your legal and subject matter expertise – not search term logistics.**
By bringing in technical experts to complement your search term development process, you are free to focus on case strategy.

Practical Steps to Develop a Defensible, Iterative Approach to Search Term Culling

Guidance from the courts, combined with pressure to control costs and minimize the risk of sanctions, have led legal teams to adopt new best practices for culling large document populations. Increasingly, they are utilizing an iterative approach to search term development, one that is driven by technical experts (including linguists and statisticians), informed and validated by attorneys, and rigorously tested and documented. Such an approach is designed to defensibly eliminate non-relevant data from review populations without sacrificing relevant data.

While the process may vary from case to case to accommodate legal teams' specific goals, timelines and resources, a typical approach to optimizing search term culling is as follows:

Step 1: Linguists collaborate with attorneys to gain an understanding of the subject matter of interest for the case. They hypothesize an extensive set of search terms and refine any search terms developed by the legal team.

Linguists consider multiple angles of assessment to account for variations in communication patterns within the specific population of interest – for example, by incorporating coverage for a range of grammatical relationships and synonyms used to express target concepts and researching jargon used by specific communities of interest.

Step 2: The hypothesized search terms are applied to the potential review population and a detailed quantitative performance report is generated describing each term's unique and aggregated contributions to the result set.

Step 3: The report is analyzed and disseminated to the case team with recommendations for statistically valid samples that should be reviewed to validate and/or refine the term set.

Step 4: Attorneys review these samples for responsiveness.

Step 5: Linguists analyze patterns of responsiveness within the reviewed samples and use that feedback to pinpoint areas of strong or weak term performance. Then, the linguists generate and test new hypotheses to arrive at a set of recommended adjustments for the search term list.

Step 6: The case team reviews the linguists' recommendations and finalizes the search term list.

Step 7: The case team undertakes a final round of statistically valid sample review to measure the quality of the search term list prior to embarking on the full review.

By utilizing an iterative process that draws upon linguistic and statistical expertise, case teams can efficiently improve the effectiveness and defensibility of their culling efforts.

Key Considerations

Some law firms and corporations are training junior attorneys and litigation support personnel to become search experts. Others are engaging the services of specialized e-discovery providers on a case-by-case basis to complement their team with expertise that is not resident in-house. Regardless of whether you bring experts in-house or collaborate with a third-party provider, the following are best practices to consider:

1. Experts should have proven track records.

Taking a crash course in statistics or linguistics - or having a long history of running searches as an IT or litigation support professional - may not qualify someone as an expert. Look for specific expertise in applying linguistics and statistics in an e-discovery context. Your experts may be called upon to testify to the robustness of the process and validity of the results.

2. Experts should be able to design and articulate a principled iterative process.

Experts should be able to design comprehensive search terms, test and refine them based on sample reviews, and measure and report on final results. The process should be flexible and adaptive, incorporating mechanisms for course-correction based on feedback from samples that have been reviewed by attorneys. This will make it possible to pinpoint and resolve issues quickly and efficiently.

How a Fortune 50 Corporation Achieved 56% Greater Reduction in Documents – and Twice as Relevant a Document Population – by Utilizing Search Expertise

A Fortune 50 corporation faced with a large document collection used search expertise to defensibly reduce the volume of its review population. The use of search expertise, sampling for attorney review and an iterative process resulted in the following:

- Overall reduction of the document collection by 80 %
- A final review population that was 56 % smaller and 50 % richer in relevant material than the results of the initial search term list
- Strong statistical data and firsthand knowledge of the quality of the review population that the outside law firm was able to use to achieve an optimal outcome in search

3. **Experts should validate the reliability and quality of the search term results.**
Experts should be able to provide you with statistically valid measurements of the reliability and quality of the final results of the search term culling effort. These measurements should provide insights into the quality of the review population that will enhance your ability to defend the culling process and negotiate search terms with opposing counsel.
4. **Experts should generate a detailed audit trail.**
Experts should provide detailed reporting and documentation for the search term culling process, including a comprehensive enumeration of what search terms were used and why. Your documentation should include a record of all key inputs, decisions and results.
5. **You should be free to focus on case strategy, not search term logistics.**
Experts should drive, manage and document a process that fits your workflow and timeline – not theirs – allowing you to think in terms of case strategy, not keyword logistics.

Conclusion

Combining legal and technical expertise with sound scientific search and information retrieval methodology promises to yield more cost-effective and defensible document review for e-discovery. It requires litigants to change their approach to discovery in some ways – engaging in more planning and design activities early in the process. Adoption of new best practices will bring tangible improvements in the quality, efficiency and cost of review.

About the Author

Amanda Jones has more than six years of experience applying advanced linguistic strategies to complex information retrieval projects, including e-discovery. In her current role as Senior Classification Analyst at Xerox Litigation Services (XLS), the e-discovery division of Xerox Corporation (NYSE: XR), Amanda oversees automated document classification and search consulting technology and services. In this role, Amanda assists corporate legal departments and law firms in formulating and validating defensible information retrieval strategies to maximize the efficacy of their document review projects. Prior to XLS, Amanda led and managed projects for the Linguistic Technology Group at H5, where she designed and supervised the implementation of linguistic models for document categorization associated with e-discovery. Amanda holds a B.A. in Linguistics from the University of Texas at Austin and an M.A. in Linguistics from the University of California, Los Angeles. Amanda can be reached at amanda.jones@xls.xerox.com.

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.

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

For more information, visit www.xerox-xls.com, call 877.273.3887 or e-mail info@xls.xerox.com.