PRISM
LITIGATION TECHNOLOGY

eDiscovery | Proportional Discovery Advisors

Don't Stop Believin':

The Staying Power of Search Term Optimization

www.prismlit.com

PRISM LITIGATION TECHNOLOGY

eDiscovery | Proportional Discovery Advisors

Don't Stop Believin':
The Staying Power of Search Term Optimization

Since the first model T rolled off of Henry Ford's assembly line, all cars have been made with a steering wheel. There is a simple reason for that: **it works.** The same can be said for keyword searches. Once the go-to tool for filtering relevant documents from non-relevant, some have criticized search terms as being limited, risky and obsolete. We disagree.

While machine learning and technology-assisted review technologies have evolved, it hardly means that a well-devised search term strategy should be abandoned. On the contrary, an innovative approach to this long-standing methodology maintains it as a viable, transparent and defensible process for companies looking to reduce the volume and expense of review. In other words, ***don't stop believin'*** in this powerful tool.

Out with the Old

The typical approach for executing a search strategy is to brainstorm a list of search terms that would be found in relevant documents, run the search terms in the dataset to be reviewed, and then separate the documents into two groups: those that do and don't contain the terms.

Unfortunately, this approach has several shortcomings:

1. Search terms are only as good as the subject matter knowledge of the people selecting them. Understanding the claims and defenses in a matter, along with the nomenclature of the vocabulary in the dataset, is essential in selecting the right language.

2. Frequently, search terms are general and may mean different things in different contexts. Simple keywords are not structured to disambiguate language, so they often produce datasets that are simultaneously overly inclusive, while still missing potentially responsive documents.

3. Limited accommodations are made for documents without searchable text or structured data like that found in spreadsheets and image file formats.

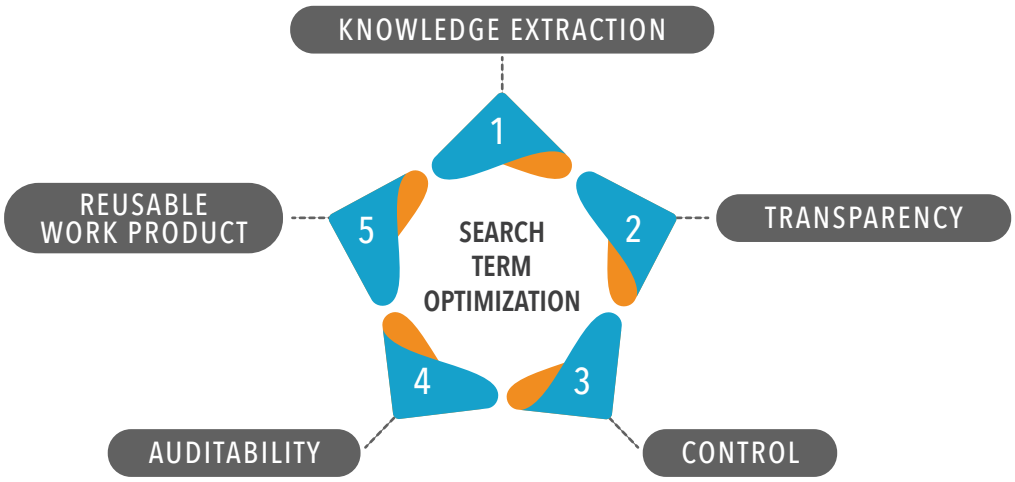
To be sure, the use of search terms requires expertise to deftly avoid the pitfalls that cause process inefficiencies; however, it is still the most widely used, transparent and defensible methodology in practice today. If deployed properly, it is a proven, cost-effective eDiscovery staple to reduce, filter and categorize large datasets.

© Prism Litigation Technology, Inc. All rights reserved.

The December 2015 revisions to the Federal Rules of Civil Procedure emphasized the acceptable practice of applying proportionality to rightsize the discovery process with the needs of the case. Counsel are required to meet and negotiate on **"any issues about disclosure, discovery, or preservation of electronically stored information, including the form or forms in which it should be produced."** [1] As such, Court rulings increasingly support cooperation, transparency and proportionality. The United States District Court ruled in ***Romero v. Allstate Insurance Co.*** that, **"Among the items about which the court expects counsel to 'reach practical agreement' without the court having to micro-manage eDiscovery are 'search terms, date ranges, key players and the like'."** [2] This new attention to reducing such an onerous burden lends itself to applying trusted techniques utilizing common, transparent and defensible tools in innovative ways.

In with the New

Certainly, steering wheels in today's vehicles have vastly improved over those found in the Model T. Similarly, the traditional search term method can be enhanced through Prism's ***Search Term Optimization*** process, which centers on five key objectives:



Prism's comprehensive narrative approach is created by merging different elements of information retrieval science, linguistics, and text and data mining analytic techniques with sound document review strategies. This approach has achieved a staggering ***80 percent reduction in data volumes.***

The core foundation for the search term narrative must revolve around the claims, defenses, and fact pattern that define the language of the who, what, where, and when of the matter. This linguistic narrative approach relies on the simple fact that if a document does not contain the defined language, then the document is unlikely to be relevant.

The narrative forms the roadmap for a comprehensive strategy designed to reduce the volume of data, while creating a thorough categorization system for organization and prioritization of review. The approach is simple, flexible and can be adapted to achieve the client's objective, whether during early case assessment, linear or technology-assisted review, or anything in between.

[1] *Federal Rules of Civil Procedure, Rule 26(f)(3)(C).*
[2] *Romero v. Allstate Ins. Co., 271 F.R.D. 96, 109 (E.D. Pa. 2010).*

© Prism Litigation Technology, Inc. All rights reserved.

PRISM LITIGATION TECHNOLOGY

eDiscovery | Proportional Discovery Advisors

Getting Started: Build the Narrative

Search term analytics is a science, best developed by experts. The first step, then, is to partner with search term optimization consultants who collaborate with the litigation team. Together, this group will read and analyze key pleadings to build a narrative of the matter. Not only does this create an understanding of the language relative to each issue of the case, but it also guides the consultants as to which issues take precedence.

The process includes the following steps:

STEP 1 **ISSUE ANALYSIS:**
Create an unambiguous definition of each issue that characterizes the claims being made and the defenses being offered.

STEP 2 **LOGICAL EXPRESSION DEFINITION:**
Define the specific expressions that encapsulate each issue. There may be multiple expressions required to convey the full meaning of the issue.

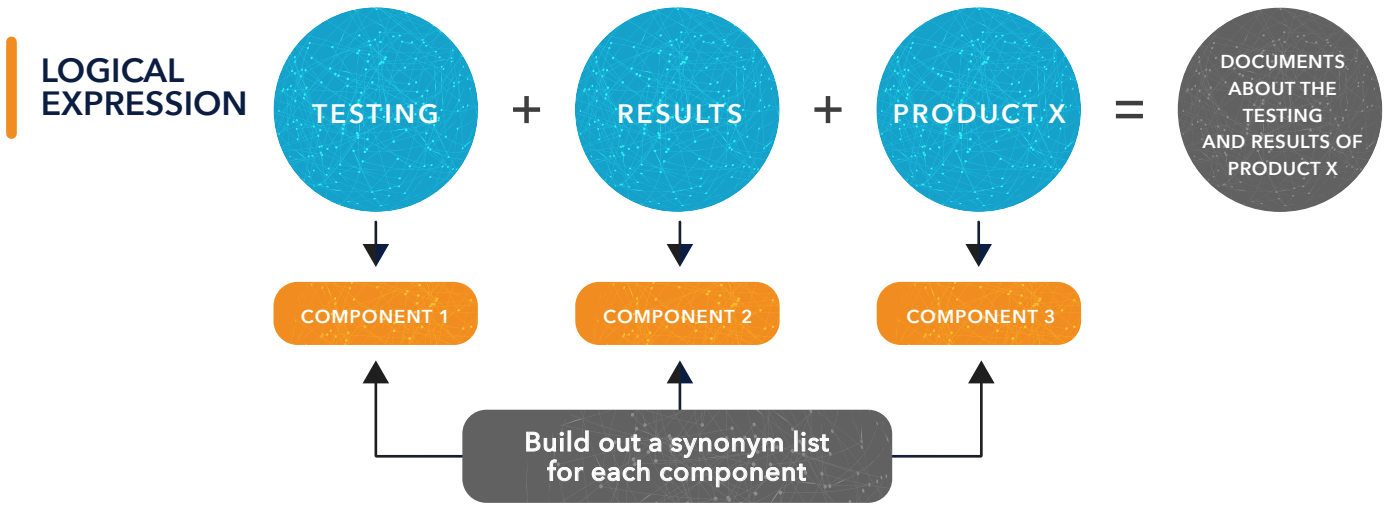
STEP 3 **COMPONENT IDENTIFICATION AND EXPANSION:**
Distill each logical expression into specific components. These components form the basis for the expansion effort, which is the identification of words that convey the same conceptual meaning (synonyms).

STEP 4 **SEARCH STRATEGIES:**
Determine the appropriate parameters to be used for proximity, as well as developing a strategy for searching non-standard, structured data, such as spreadsheet or database files.

The objective is to find the documents that are about each issue or, in other words, that ***contain language*** that makes the document relevant.

For example, assume the matter is a contractual dispute with the plaintiff suing for breach of contract. One issue involves the performance failure of a product that was ordered and delivered under the terms of the contract. The specific item of interest is the results of quality control testing of Product X. The logical expression definition then becomes finding documents specifically related to the ***testing*** and the ***results*** for ***Product X***. Logically, for a document to be relevant to this concept, it would need to contain language for each of the three components.

© Prism Litigation Technology, Inc. All rights reserved.



But it is not that simple. Recognizing that custodians may not use the term ***testing*** in their documents or email exchanges, instead calling it quality control, trials or other synonyms, the consulting team applies both its own expertise and the legal team's input to expand each of the components to reflect any synonyms that are contained within the vocabulary of the dataset. The same process is performed for the components of ***results*** and ***Product X***.

Once this assessment is completed, the consulting team builds the corresponding search syntax and appropriate Boolean connectors to form a structured query designed to identify the documents that are related to each logical expression. The goal is twofold: to provide granular issue identification of the documents, while achieving optimal recall and precision in the results.

The Rhythm of Precision and Recall

Balancing precision and recall form the next challenge in constructing the narrative. Anchored in the information retrieval world, precision and recall are used to measure the effectiveness of a single search or a group of searches to identify relevant documents.

> ***Recall is the fraction / percentage of relevant documents that were retrieved.***

> ***Precision is the fraction / percentage of retrieved documents that are actually relevant.***

The goal in any information retrieval exercise is high recall ***and*** precision, or in other words, to find all relevant documents and ensure that the majority of these documents contain information connected to the case, while also reducing the retrieval of false positives. These two goals often work at odds with one another. While casting a wide net ensures that everything is captured, it can also result in high false positive results and low precision.

The legal team must make the strategic decision as to which measure is more critical. It may be adequate to find only a few specific instances of documents relevant to an issue or it may be important to find every instance within the corpus. In the previous example, if the consulting team only searched for the word ***testing***, it would result in very high recall and very low precision.

© Prism Litigation Technology, Inc. All rights reserved.

This is because many of those documents would likely be about topics other than the testing results for Product X.

Precision can be improved, however, by adding context or multiple components. For example, consider if the consultants searched for ***testing*** and ***Product X*** together. While this would yield better results, many documents could be retrieved that do not discuss the actual results of the test, a critical component of the narrative.

Creating a search that combines all three components of the narrative will result in the highest precision, while minimizing false recall.

This narrative approach allows for an easy increase in recall or precision by adding or removing components or adjusting the proximity. Understanding the delicate balance of recall and precision requires expertise and the strategic use of random sampling to evaluate search results and adjust accordingly.

Leveraging Search Term Narratives

One of the shortcomings of many search term strategies is that litigation teams tend to use them and lose them. Time and effort is expended to build the list of words and run the searches, but the knowledge that is gleaned relative to the documents retrieved is often lost.

The consulting team can use the narrative and search lexicon to compile a completely categorized corpus, organizing and assisting in prioritizing document review. Knowledge gained should be leveraged for ongoing case needs including future productions, along with deposition and trial preparation. In addition, for corporations with pattern litigation, the work product that is created can easily be repurposed with minimal adjustment for similar matters.

Leverage search term work product post-production, to assist in discovery and trial preparation.

Build a corporate vocabulary that can be enhanced and utilized across other litigation.

Conclusion

We have not stopped believing in the steering wheel, nor should we abandon the use of search terms as a valuable process in the eDiscovery toolbox. Over time, litigation teams have seen multiple iterations of artificial intelligence, technology-assisted review, and other tools ebb and flow in their utility and practicality.

Despite these enhancements, search term development remains as fundamental to eDiscovery as the steering wheel is to a vehicle.

Led by an experienced consulting team, Prism's ***Search Term Optimization*** methodology enriches common search terms by adding layers of linguistic and data science expertise to create a fully defensible, transparent, and cogent workflow. This approach, when paired with leveraging proportionality to rightsize discovery, can steer you efficiently into practicing in the 21st century.

© Prism Litigation Technology, Inc. All rights reserved.